**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ**



**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «ЛИПЕЦКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Институт компьютерных наук

Кафедра автоматизированных систем управления

ЛАБОРАТОРНАЯ РАБОТА №2

по дисциплине «Машинное обучение»

«Классификация данных»

Студенты М-РИТ-25 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Киселев М.С.

(подпись, дата) Красиков И.А

Руководитель  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Сараев П.В.

(подпись, дата)

Липецк 2025

**Цель работы**

Изучение алгоритмов построения классификаторов, оценки качества классификации и средств классификации данных в языке Python.

**Ход работы**

Исходный код программы в приложении 1.

1. Вычисление различных величин с помощью средств языка Python

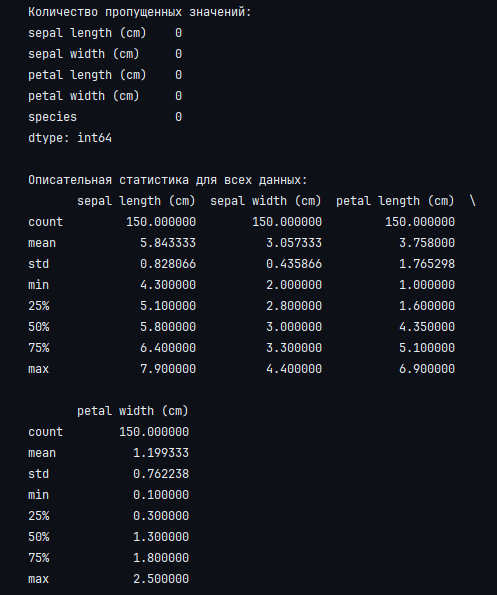


Рисунок 1 – Вывод в консоль описательной статистики по всем данным

1. Сводная информация об обучающем множестве

\

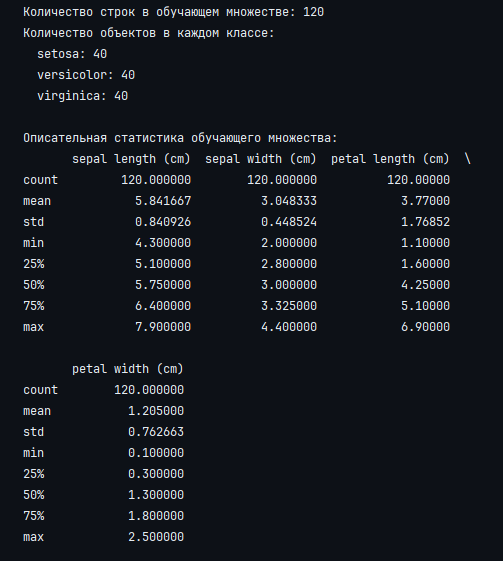


Рисунок 2 – Вывод в консоль описательной статистики по обучающему множеству

1. **НАИВНЫЙ БАЙЕСОВСКИЙ КЛАССИФИКАТОР:**

==================================================

РЕЗУЛЬТАТЫ ДЛЯ 4 ПАРАМЕТРОВ:

Точность классификации: 0.967

Уровень ошибки: 0.033

По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 1.000

Чувствительность (Recall): 0.900

Специфичность (Specificity): 1.000

virginica:

Точность (Precision): 0.909

Чувствительность (Recall): 1.000

Специфичность (Specificity): 0.950

Матрица ошибок:

[[10 0 0]

[ 0 9 1]

[ 0 0 10]]

Наивный Байес (4 параметра) - ОБУЧАЮЩЕЕ МНОЖЕСТВО:

Точность классификации: 0.958

Уровень ошибки: 0.042

По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 0.927

Чувствительность (Recall): 0.950

Специфичность (Specificity): 0.963

virginica:

Точность (Precision): 0.949

Чувствительность (Recall): 0.925

Специфичность (Specificity): 0.975

Матрица ошибок:

[[40 0 0]

[ 0 38 2]

[ 0 3 37]]

РЕЗУЛЬТАТЫ ДЛЯ 2 ПАРАМЕТРОВ (sepal length и width):

Точность классификации: 0.700

Уровень ошибки: 0.300

По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 0.556

Чувствительность (Recall): 0.500

Специфичность (Specificity): 0.800

virginica:

Точность (Precision): 0.545

Чувствительность (Recall): 0.600

Специфичность (Specificity): 0.750

Матрица ошибок:

[[10 0 0]

[ 0 5 5]

[ 0 4 6]]

Наивный Байес (2 параметра) - ОБУЧАЮЩЕЕ МНОЖЕСТВО:

Точность классификации: 0.808

Уровень ошибки: 0.192

По классам:

setosa:

Точность (Precision): 0.976

Чувствительность (Recall): 1.000

Специфичность (Specificity): 0.988

versicolor:

Точность (Precision): 0.696

Чувствительность (Recall): 0.800

Специфичность (Specificity): 0.825

virginica:

Точность (Precision): 0.758

Чувствительность (Recall): 0.625

Специфичность (Specificity): 0.900

Матрица ошибок:

[[40 0 0]

[ 0 32 8]

[ 1 14 25]]

1. **ДЕРЕВЬЯ РЕШЕНИЙ:**

==================================================

РЕЗУЛЬТАТЫ ДЛЯ 4 ПАРАМЕТРОВ:

Точность классификации: 0.933

Уровень ошибки: 0.067

По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 0.900

Чувствительность (Recall): 0.900

Специфичность (Specificity): 0.950

virginica:

Точность (Precision): 0.900

Чувствительность (Recall): 0.900

Специфичность (Specificity): 0.950

Матрица ошибок:

[[10 0 0]

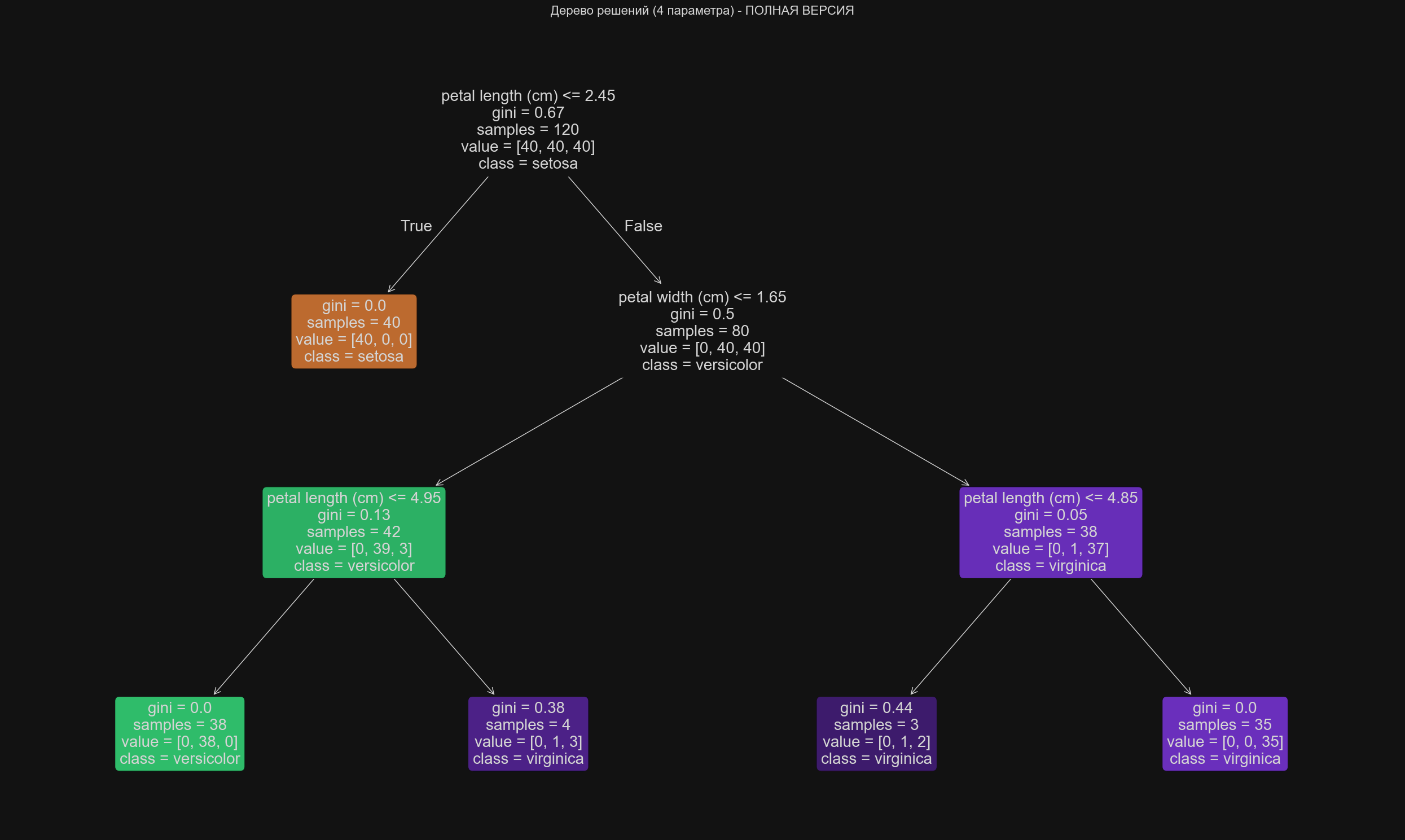
[ 0 9 1]

[ 0 1 9]]

Дерево решений (4 параметра) - ОБУЧАЮЩЕЕ МНОЖЕСТВО:

Точность классификации: 1.000

Уровень ошибки: 0.000



По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

virginica:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

Матрица ошибок:

[[40 0 0]

[ 0 40 0]

[ 0 0 40]]

Глубина дерева (4 параметра): 5

Количество листьев (4 параметра): 8

РЕЗУЛЬТАТЫ ДЛЯ 2 ПАРАМЕТРОВ:

Точность классификации: 0.567

Уровень ошибки: 0.433

По классам:

setosa:

Точность (Precision): 0.909

Чувствительность (Recall): 1.000

Специфичность (Specificity): 0.950

versicolor:

Точность (Precision): 0.286

Чувствительность (Recall): 0.200

Специфичность (Specificity): 0.750

virginica:

Точность (Precision): 0.417

Чувствительность (Recall): 0.500

Специфичность (Specificity): 0.650

Матрица ошибок:

[[10 0 0]

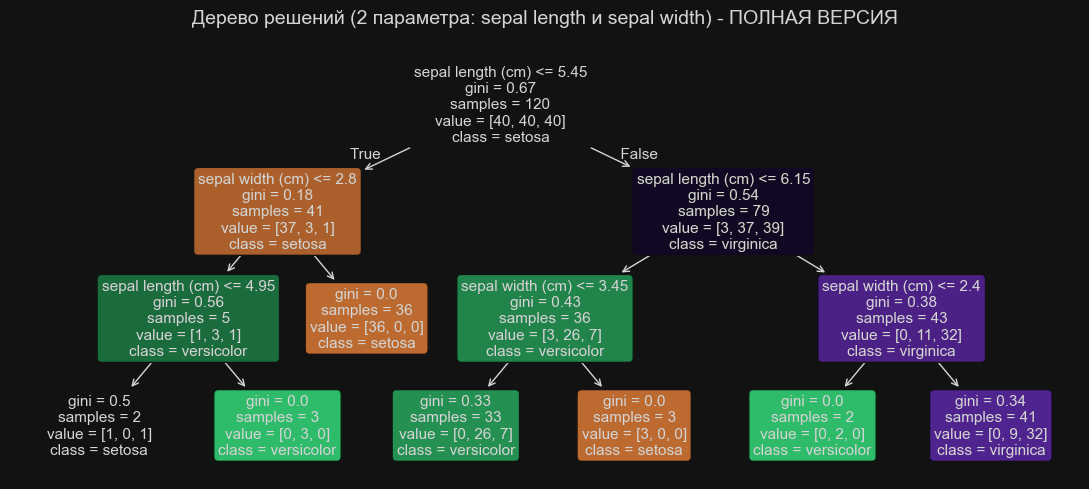
[ 1 2 7]

[ 0 5 5]]

Дерево решений (2 параметра) - ОБУЧАЮЩЕЕ МНОЖЕСТВО:

Точность классификации: 0.950

Уровень ошибки: 0.050



По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 0.870

Чувствительность (Recall): 1.000

Специфичность (Specificity): 0.925

virginica:

Точность (Precision): 1.000

Чувствительность (Recall): 0.850

Специфичность (Specificity): 1.000

Матрица ошибок:

[[40 0 0]

[ 0 40 0]

[ 0 6 34]]

Глубина дерева (2 параметра): 11

Количество листьев (2 параметра): 33

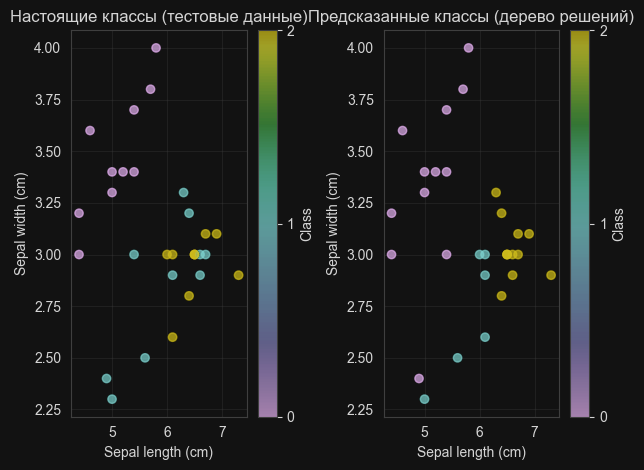


Рисунок 3 – Графическое сравнение результатов для двух параметров

для тестового множества

1. **МЕТОД ОПОРНЫХ ВЕКТОРОВ (SVM)**

РЕЗУЛЬТАТЫ ДЛЯ 4 ПАРАМЕТРОВ:

Точность классификации: 0.967

Уровень ошибки: 0.033

По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 1.000

Чувствительность (Recall): 0.900

Специфичность (Specificity): 1.000

virginica:

Точность (Precision): 0.909

Чувствительность (Recall): 1.000

Специфичность (Specificity): 0.950

Матрица ошибок:

[[10 0 0]

[ 0 9 1]

[ 0 0 10]]

SVM (4 параметра) - ОБУЧАЮЩЕЕ МНОЖЕСТВО:

Точность классификации: 0.983

Уровень ошибки: 0.017

По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 0.975

Чувствительность (Recall): 0.975

Специфичность (Specificity): 0.988

virginica:

Точность (Precision): 0.975

Чувствительность (Recall): 0.975

Специфичность (Specificity): 0.988

Матрица ошибок:

[[40 0 0]

[ 0 39 1]

[ 0 1 39]]

РЕЗУЛЬТАТЫ ДЛЯ 2 ПАРАМЕТРОВ:

Точность классификации: 0.700

Уровень ошибки: 0.300

По классам:

setosa:

Точность (Precision): 1.000

Чувствительность (Recall): 1.000

Специфичность (Specificity): 1.000

versicolor:

Точность (Precision): 0.556

Чувствительность (Recall): 0.500

Специфичность (Specificity): 0.800

virginica:

Точность (Precision): 0.545

Чувствительность (Recall): 0.600

Специфичность (Specificity): 0.750

Матрица ошибок:

[[10 0 0]

[ 0 5 5]

[ 0 4 6]]

SVM (2 параметра) - ОБУЧАЮЩЕЕ МНОЖЕСТВО:

Точность классификации: 0.817

Уровень ошибки: 0.183

По классам:

setosa:

Точность (Precision): 0.976

Чувствительность (Recall): 1.000

Специфичность (Specificity): 0.988

versicolor:

Точность (Precision): 0.702

Чувствительность (Recall): 0.825

Специфичность (Specificity): 0.825

virginica:

Точность (Precision): 0.781

Чувствительность (Recall): 0.625

Специфичность (Specificity): 0.912

Матрица ошибок:

[[40 0 0]

[ 0 33 7]

[ 1 14 25]]

1. СРАВНЕНИЕ РЕЗУЛЬТАТОВ.



Рисунок 4 – Сравнение методов

Заключение

Лучший метод для 4 параметров: Наивный Байес (точность: 0.967). Лучший метод для 2 параметров: Наивный Байес (точность: 0.700). Использование 4 параметров дает лучшую точность классификации.