

Министерство образования Республики Беларусь  
Учреждение образования  
«Брестский Государственный технический университет»  
Кафедра ИИТ

Лабораторная работа №3  
По дисциплине «Основы машинного обучения»  
Тема: **«Сравнение классических методов классификации»**

Выполнил:  
Студент 3 курса  
Группы АС-66  
Гончерёнок К.А.  
Проверил:  
Крощенко А. А.

Брест 2025

Цель работы: На практике сравнить работу нескольких алгоритмов классификации, таких как метод k-ближайших соседей (k-NN), деревья решений и метод опорных векторов (SVM). Научиться подбирать гиперпараметры моделей и оценивать их влияние на результат.

Вариант 2  
Ход работы:

Задание:

- Breast Cancer Wisconsin
- Определить, является ли опухоль злокачественной (malignant) или доброкачественной (benign);
- Задания:
  1. Загрузите данные и выполните стандартизацию признаков;
  2. Разделите данные на обучающую и тестовую части;
  3. Обучите три классификатора: k-NN, Decision Tree и SVM;
  4. Для каждой модели постройте матрицу ошибок и рассчитайте метрики precision, recall и F1-score для класса "злокачественная опухоль";
  5. Сравните модели и укажите, какая из них наиболее надежна для минимизации ложноотрицательных прогнозов (когда злокачественная опухоль определяется как доброкачественная).

```
print("ИССЛЕДОВАНИЕ k-NN С РАЗНЫМ K:")
```

```
print(f"Лучшая точность: k={best_k_accuracy}, точность={best_accuracy:.4f}")
print(f"Меньше всего ложноотрицательных: k={best_k_fn}, FN={best_fn}")
```

```
print("\nМЕТРИКИ ДЛЯ ЗЛОКАЧЕСТВЕННЫХ ОПУХОЛЕЙ:")
```

```
print(f"Матрица ошибок:\n{cm}")
```

```
print(f"Ложноотрицательные: {cm[1,0]}")
```

```
print("\nСРАВНЕНИЕ МОДЕЛЕЙ:")
```

```
print(f"{name}: {fn} ложноотрицательных")
```

```
print(f"\nЛУЧШАЯ МОДЕЛЬ: {best_model[0]} ({best_model[1]} FN)")
```

```

Размер признаков: (569, 30)
Обучающая выборка: (398, 30)
Тестовая выборка: (171, 30)
Обучение k-NN...
Обучение Decision Tree...
Обучение SVM...
k-NN - Отчет по классификации:
      precision    recall  f1-score   support

Доброкачественная      0.95      1.00      0.97      107
Злокачественная       1.00      0.91      0.95       64

      accuracy      0.96      171
      macro avg      0.97      0.95      0.96      171
      weighted avg      0.97      0.96      0.96      171

Decision Tree - Отчет по классификации:
      precision    recall  f1-score   support

Доброкачественная      0.91      0.93      0.92      107
Злокачественная       0.89      0.84      0.86       64

      accuracy      0.90      171
      macro avg      0.90      0.89      0.89      171
      weighted avg      0.90      0.90      0.90      171

SVM - Отчет по классификации:
      precision    recall  f1-score   support

Доброкачественная      0.95      1.00      0.97      107
Злокачественная       1.00      0.91      0.95       64

      accuracy      0.96      171
      macro avg      0.97      0.95      0.96      171
      weighted avg      0.97      0.96      0.96      171

Сравнение метрик для класса 'Злокачественная':
      precision    recall  f1_score
k-NN          1.000000  0.90625  0.95082
Decision Tree  0.885246  0.84375  0.86400
SVM           1.000000  0.90625  0.95082
Лучшая модель для минимизации ложноотрицательных: k-NN
Полнота: 0.9062

Анализ ложноотрицательных прогнозов:
k-NN: 6 ложноотрицательных (9.38%)
Decision Tree: 10 ложноотрицательных (15.62%)
SVM: 6 ложноотрицательных (9.38%)

Рекомендуемая модель: k-NN
Полнота: 0.9062

```

Вывод: На практике сравнил работу нескольких алгоритмов классификации, таких как метод k-ближайших соседей (k-NN), деревья решений и метод опорных векторов (SVM). Научился подбирать гиперпараметры моделей и оценивать их влияние на результат.

