

**Московский авиационный институт  
(Национальный исследовательский университет)**

Институт: «Информационные технологии и прикладная математика»  
Кафедра: 806 «Вычислительная математика и программирование»

**Лабораторная работа № 2  
по курсу «Искусственный интеллект»**

Студент:	Обыденкова Ю. Ю.
Группа:	М8О-308Б-18
Оценка:	
Дата:	

Москва, 2021

Постановка задачи:

Необходимо реализовать алгоритмы машинного обучения. Применить данные алгоритмы на наборы данных, подготовленных в первой лабораторной работе. Провести анализ полученных моделей, вычислить метрики классификатора. Произвести тюнинг параметров в случае необходимости. Сравнить полученные результаты с моделями реализованными в `scikit-learn`. Аналогично построить метрики классификации. Показать, что полученные модели не переобучились. Также необходимо сделать выводы о применимости данных моделей к вашей задаче.

Датасет: статистика преступлений в РФ за последнее время.

Логистическая регрессия:

Это статическая модель, которая используется для прогнозирования вероятности возникновения некоторого события путем его сравнения с логистической кривой.

Результаты:

```

sklearn log:
[0 1 0 0 0 0 0 1 0 0 1 0 1 1 1 1 1 1 1 1 1]
accur: 0.8695652173913043
      precision    recall  f1-score   support

         0         0.80      0.89      0.84            9
         1         0.92      0.86      0.89           14

   accuracy              0.87            23
  macro avg              0.86      0.87      0.87            23
weighted avg              0.87      0.87      0.87            23

my log:
[0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
my accur: 0.8695652173913043
      precision    recall  f1-score   support

         0         0.80      0.89      0.84            9
         1         0.92      0.86      0.89           14

   accuracy              0.87            23
  macro avg              0.86      0.87      0.87            23
weighted avg              0.87      0.87      0.87            23

```

Точности алгоритмов и sklearn совпадают.

Дерево решений:

Это жадное, нисходящее рекурсивное разбиение. Энтропия – мера случайности или неопределенности, уровень ее от 0 до 1.

Результаты:

```
sklearn Dtree:
[0 0 0 0 0 0 0 1 1 0 0 1 1 1 1 1 1 1 1 1]
accur:
0.9047619047619048
      precision    recall  f1-score   support

     0       0.82      1.00      0.90         9
     1       1.00      0.83      0.91        12

   accuracy          0.90         21
  macro avg       0.91      0.92      0.90         21
weighted avg       0.92      0.90      0.91         21

my Dtree:
[0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1]
my accur: 0.9047619047619048
      precision    recall  f1-score   support

     0       0.82      1.00      0.90         9
     1       1.00      0.83      0.91        12

   accuracy          0.90         21
  macro avg       0.91      0.92      0.90         21
weighted avg       0.92      0.90      0.91         21
```

Точности алгоритмов и sklearn совпадают.

Случайный лес:

Это множество решающих деревьев. В задаче регрессии их ответы усредняются, в задаче классификации принимается решение голосованием пр большинству.

Результаты:

```
sklearn RF:
[0 0 0 0 1 0 0 0 0 1 1 1 1 1]
accur:
0.8571428571428571
      precision    recall  f1-score   support

     0       0.88      0.88      0.88         8
     1       0.83      0.83      0.83         6

   accuracy          0.86         14
  macro avg          0.85         14
weighted avg          0.86         14

my RF:
[0 0 0 0 1 0 0 0 0 1 1 1 1 1]
my accur: 0.8571428571428571
      precision    recall  f1-score   support

     0.0       0.88      0.88      0.88         8
     1.0       0.83      0.83      0.83         6

   accuracy          0.86         14
  macro avg          0.85         14
weighted avg          0.86         14
```

Точности алгоритмов и sklearn совпадают.

```
LR train: 0.8846153846153846
DT train: 0.9538461538461539
RF train: 0.9307692307692308
```

Модели не переобучились, тк точность несильно упала.