

Искусственный интеллект в маленьком городе

Автор: Лаптева Арина, 10 «Б» класс, НМБОУ «Гимназия
№11»

Научный руководитель: Мангазеев А.Ю., учитель
информатики и ИКТ НМБОУ «Гимназии №11»

Цель

Нашей целью стало показать сферы применения алгоритмов искусственного интеллекта на языке программирования Python с помощью различных библиотек машинного обучения.

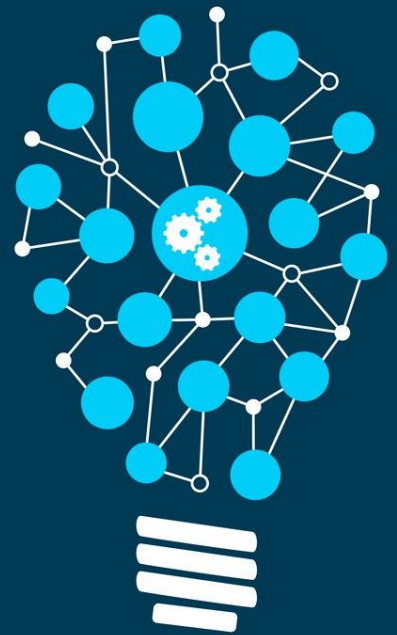
Задачи и актуальность исследования

1. Изучить синтаксис библиотек Pandas, Matplotlib и SKLearn.
2. Обучить программу на основе многолетних данных о погоде нашего региона, составить предсказания и вычислить процент ошибки такого метода предсказаний.
3. Изучить алгоритмы принятия решений искусственным интеллектом на примере кредитного скоринга.
4. С помощью библиотек машинного обучения составить наглядную аналитику ЕГЭ по русскому языку и математике.
5. Сделать вывод о сферах применения искусственного интеллекта и поставить дальнейшие цели для его изучения.

Инструменты для работы с алгоритмами машинного обучения

- ▶ *Контролируемое обучение.* Компьютер представлен с примерными вводами и желаемыми выводами. Для удобства и наглядности работы мы использовали *Jupyter Notebook* - интерактивную среду для запуска программного кода в браузере

MACHINE
LEARNING



Инструменты для работы с алгоритмами машинного обучения

► Для работы с данными и обучения алгоритмов мы использовали следующие библиотеки:

1. *Pandas* - библиотека Python для обработки и анализа данных.
2. *Matplotlib* — библиотека на языке программирования Python для визуализации данных двумерной и трёхмерной графикой.
3. *Scikit-learn* - очень популярный инструмент и самая известная библиотека Python для машинного обучения.

Предсказания искусственного интеллекта

Формирование обучения алгоритма на примере погодных данных

Теперь мы должны сформировать матрицу обучения(X) и вектор ответов(y)

Объектом классификатора станет решающее дерево DecisionTreeClassifier из модуля tree. В скобках указывается глубина чтения данных для обучения

```
In [27]: X = new_data.loc[:, ['temperature', 'pressure', 'humidity']]  
y = new_data['rainfall']
```

```
In [28]: classifier = tree.DecisionTreeClassifier(max_depth = 30)  
#метод fit обучает алгоритм по обучающей выборке  
classifier.fit(X, y)
```

```
Out[28]: DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=30,  
                                max_features=None, max_leaf_nodes=None,  
                                min_impurity_decrease=0.0, min_impurity_split=None,  
                                min_samples_leaf=1, min_samples_split=2,  
                                min_weight_fraction_leaf=0.0, presort=False, random_state=None,  
                                splitter='best')
```

Алгоритм обучения сформирован, теперь осталось проверить его предсказания

```
In [39]: #метод predict формирует предсказание на основе матрицы обучения  
prediction = classifier.predict([[20, 747, 89]])  
print(prediction)  
  
['clean']
```

Предсказания искусственного интеллекта

Выявление ошибки алгоритма предсказания осадков

Алгоритм выдаёт нам предсказания. Но какой процент ошибки в них? Это мы можем выяснить с помощью метода k ближайших соседей(kNN)

```
In [41]: from sklearn.neighbors import KNeighborsClassifier  
import numpy as np
```

```
In [42]: knn = KNeighborsClassifier(n_neighbors=3)  
knn.fit(X, y)
```

```
Out[42]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',  
metric_params=None, n_jobs=1, n_neighbors=3, p=2,  
weights='uniform')
```

После того, как модель обучена, мы можем предсказывать значение целевого признака по входным признакам для новых объектов. Делается это с помощью метода predict.

Как мы можем видеть ошибка составляет 18.4% у обучающей выборки.

```
In [43]: y_predict = knn.predict(X)  
err = np.mean(y != y_predict) #метод mean библиотеки numpy определяет среднее арифметическое по заданному вектору  
print(err)
```

```
0.1842410196987254
```

Аналитика ЕГЭ и машинное обучение

Считывание данных в формате Excel из файлов

Аналитика ЕГЭ по математике и русскому языку

Считывание данных для последующей работы

```
In [1]: import pandas as pd  
import os
```

```
In [3]: all_files_EGE = os.listdir()  
files = ['./EGE and OGE/16-17/результаты ЕГЭ 2017/' + i for i in os.listdir('./EGE and OGE/16-17/результаты ЕГЭ 2017') if 'xls' in
```

```
In [4]: for i in files:  
    print(i)
```

```
./EGE and OGE/16-17/результаты ЕГЭ 2017/англ.яз.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/Биология.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/география.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/информатика.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/История 1.xlsx  
./EGE and OGE/16-17/результаты ЕГЭ 2017/История 2.xlsx  
./EGE and OGE/16-17/результаты ЕГЭ 2017/литература.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/Литература2.xlsx  
./EGE and OGE/16-17/результаты ЕГЭ 2017/матем. профиль.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/математика база.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/общество.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/рус.яз.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/физика.xls  
./EGE and OGE/16-17/результаты ЕГЭ 2017/Химия.xlsx
```

```
In [11]: rus = pd.read_excel(files[11].strip(), header = 6)  
rus.drop(rus.columns[[0, 1,3,5,7,9,11,15,17,19,21,23,25]], axis=1, inplace=True)  
rus = rus[:33]  
rus
```


Аналитика ЕГЭ и машинное обучение

Максимальные и минимальные баллы по обоим предметам

Математика

```
In [27]: math.iloc[math['Балл'].idxmax()]
```

```
Out[27]: Код МСУ          901
Код ОО          404004
Класс          11А
Код ППЭ          801
Аудитория        17
Фамилия          Иванова
Имя             Наталья
Отчество         Андреевна
Серия           3213
Номер           281946
Задания с кратким ответом      ++++++++
Задания с развернутым ответом  2(2)0(2)2(2)0(3)3(3)0(4)1(4)
Первичный балл      20
Балл             82
Name: 59, dtype: object
```

```
In [28]: math.iloc[math['Балл'].idxmin()]
```

```
Out[28]: Код МСУ          901
Код ОО          404004
Класс          11А
Код ППЭ          801
Аудитория        15
Фамилия          Подлипская
Имя             Яна
Отчество         Андреевна
Серия           3213
Номер           332174
Задания с кратким ответом      -+-----+
Задания с развернутым ответом  0(2)0(2)0(2)0(3)0(3)0(4)0(4)
Первичный балл      4
Балл             18
Name: 7, dtype: object
```

Русский язык

```
In [25]: rus.iloc[rus['Балл'].idxmax()]
```

```
Out[25]: Код МСУ          901
Код ОО          404004
Класс          11В
Код ППЭ          801
Аудитория        14
Фамилия          Паскаль
Имя             Анна
Отчество         Александровна
Серия           3213
Номер           331939
Задания с кратким ответом      2+++++5++++++2++++++4
Задания с развернутым ответом  1(1)3(3)1(1)3(3)2(2)2(2)3(3)3(3)2(2)1(2)1(1)1(1)
Первичный балл      56
Балл             98
Name: 1, dtype: object
```

```
► In [26]: rus.iloc[rus['Балл'].idxmin()]
```

```
Out[26]: Код МСУ          901
Код ОО          404004
Класс          11В
Код ППЭ          801
Аудитория        13
Фамилия          Чудов
Имя             Кирилл
Отчество         Сергеевич
Серия           3213
Номер           281324
Задания с кратким ответом      2++++-4++++++2-----+2
Задания с развернутым ответом  1(1)1(3)1(1)3(3)2(2)2(2)1(3)2(3)2(2)2(2)1(1)1(1)
Первичный балл      41
Балл             67
Name: 22, dtype: object
```

Аналитика ЕГЭ и машинное обучение

Количество высокобалльников

Русский язык

```
In [16]: countMax_rus = 0
for i in rus['Балл']:
    if int(i) > 80:
        countMax_rus += 1
countMax_rus
```

Out[16]: 17

Математика

```
In [17]: countMax_math = 0
for i in math['Балл']:
    if int(i) > 80:
        countMax_math += 1
countMax_math
```

Out[17]: 1

Количество учеников не прошедших порог

Русский язык

```
In [14]: count = 0
min_ball = 36
for i in rus['Балл']:
    if int(i) < min_ball:
        count += 1
count
```

Out[14]: 0

Математика

```
In [15]: count1 = 0
min_ball = 27
for i in math['Балл']:
    if int(i) < min_ball:
        count1 += 1
count1
```

Out[15]: 5

Аналитика ЕГЭ и машинное обучение

Анализ первой части на примере ЕГЭ по русскому языку

Анализ первой части ЕГЭ

```
In [18]: #функция перевода "+" и "-" в 1 и 0
def signsInDigits(s):
    new_s = ''
    for i in s:
        if i == '+':
            new_s += '1'
        elif i == '-':
            new_s += '0'
        else:
            new_s += i
    A = [int(i) for i in new_s]
    return A
```

```
In [34]: #добавляем сумму первой части ЕГЭ по русскому языку
summs = []
for row in new_rus.itertuples():
    summa = 0
    for j in range(2, len(row)):
        summa += int(row[j])
    summs.append(summa)

new_rus.insert(25, 'summa', summs)
new_rus
```

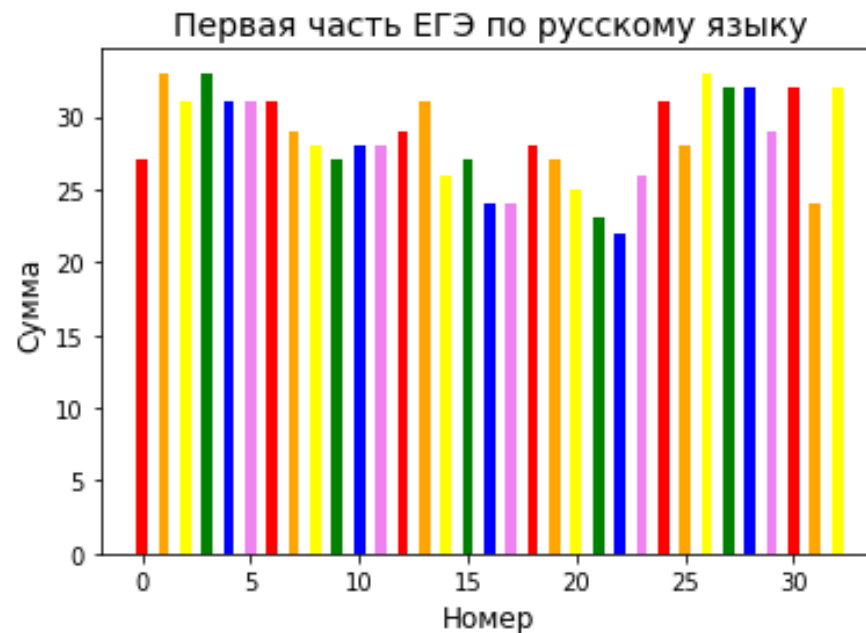
```
In [59]: #создаём новые DataFrame для наглядности
new_rus = pd.DataFrame(columns = [str(i+1) for i in range(24)])
new_math = pd.DataFrame(columns = [str(i+1) for i in range(12)])
```

```
In [46]: #переводим значения для русского языка
for i, j in enumerate(rus['Задания с кратким ответом']):
    digits = signsInDigits(j)
    new_rus.loc[i] = [k for k in digits]

new_rus.insert(0, 'Фамилия', rus['Фамилия'])
new_rus
```

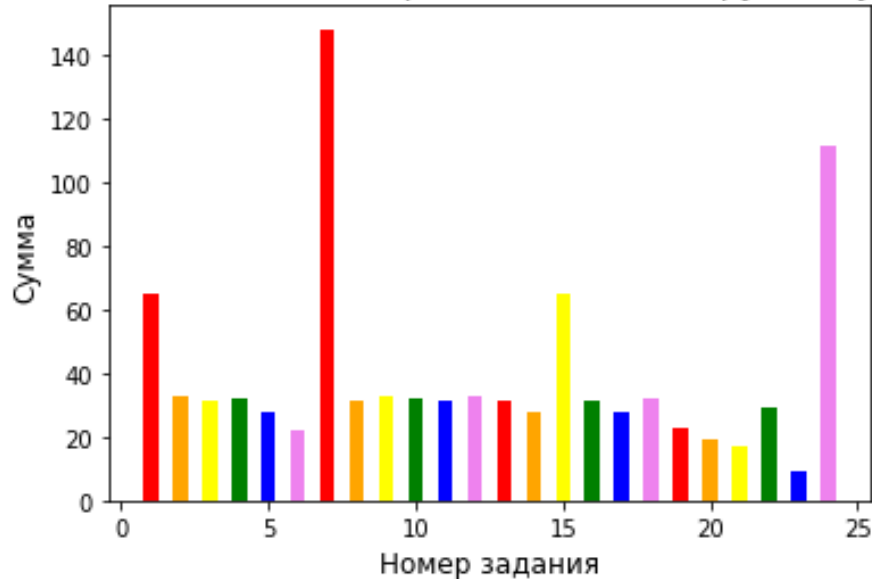
```
#добавляем сумму по заданиям первой части ЕГЭ по русскому языку
sum_rus = []
for i in range(1,25):
    sum_rus.append(new_rus[str(i)].sum())
new_rus.loc[33] = ['Сумма по заданиям'] + [j for j in sum_rus] + [' ']
new_rus
```

```
In [38]: colorsr = ['red', 'orange', 'yellow', 'green', 'blue', 'violet']
import matplotlib.pyplot as plt
num = [i for i in range(33)]
sums = [int(new_rus['summa'][i]) for i in range(len(new_rus['summa'])-1)]
plt.bar(num, sums, width = 0.5, color = colorsr)
plt.title("Первая часть ЕГЭ по русскому языку", fontsize=14)
plt.xlabel("Номер", fontsize=12)
plt.ylabel("Сумма", fontsize=12)
plt.show()
```



```
In [40]: num = [i for i in range(1, 25)]  
         sums = [int(new_rus.loc[33][i]) for i in range(1, 25)]  
         plt.bar(num, sums, width = 0.5, color = colorsr)  
         plt.title("Выполнение заданий первой части ЕГЭ по русскому языку", fontsize=14)  
         plt.xlabel("Номер задания", fontsize=12)  
         plt.ylabel("Сумма", fontsize=12)  
         plt.show()
```

Выполнение заданий первой части ЕГЭ по русскому языку



Аналитика ЕГЭ и машинное обучение

Анализ второй части на примере ЕГЭ по математике

Математика

```
In [46]: #разбиваем строчку на разные ячейки
math_2 = pd.DataFrame(columns = [str(i+1) for i in range(7)])
for i, j in enumerate(math['Задания с развёрнутым ответом']):
    digits2 = points(j)
    math_2.loc[i] = [k for k in digits2]

math_2.insert(0, 'Фамилия', math['Фамилия'][:])

#подсчитываем сумму баллов второй части у каждого выпускника
summs2 = []
for row in math_2.itertuples():
    summa = 0
    for j in range(2, len(row)):
        summa += int(row[j])
    summs2.append(summa)

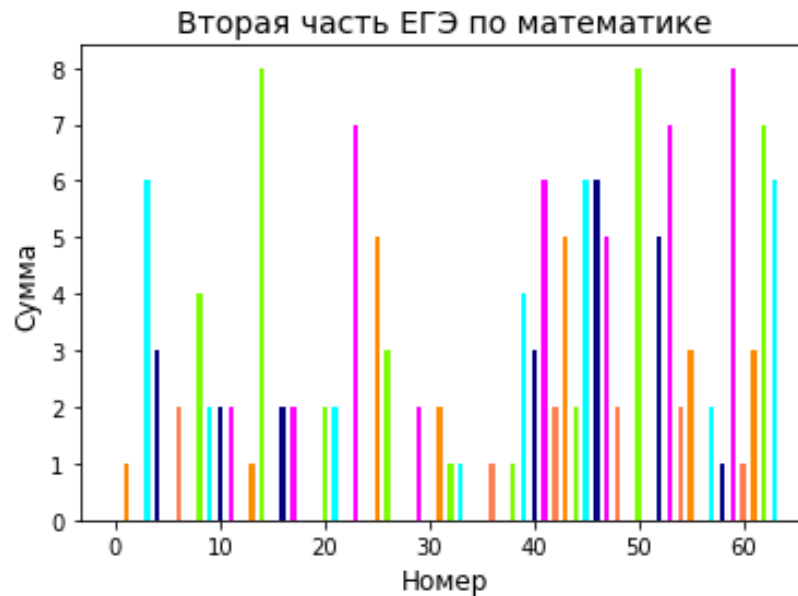
math_2.insert(8, 'summa', summs2)

#подсчитываем сумму по заданиям второй части
sum_math2 = []
for i in range(1,8):
    sum_math2.append(math_2[str(i)].sum())
math_2.loc[64] = ['Сумма по заданиям'] + [j for j in sum_math2] + [' ']
```

Визуализация второй части ЕГЭ по математике

```
In [61]: colorsm = ['coral', 'darkorange', 'lawngreen', 'aqua', 'navy', 'fuchsia']
```

```
In [62]: num = [i for i in range(64)]  
sums = [int(math_2['summa'][i]) for i in range(len(math_2['summa'])-1)]  
plt.bar(num, sums, width = 0.5, color = colorsm)  
plt.title("Вторая часть ЕГЭ по математике", fontsize=14)  
plt.xlabel("Номер", fontsize=12)  
plt.ylabel("Сумма", fontsize=12)  
plt.show()
```



```
In [63]: num = [i for i in range(1, 8)]  
         sums = [int(math_2.loc[64][i]) for i in range(1, 8)]  
         plt.bar(num, sums, width = 0.5, color = colorsm)  
         plt.title("Выполнение заданий второй части ЕГЭ по математике", fontsize=14)  
         plt.xlabel("Номер задания", fontsize=12)  
         plt.ylabel("Сумма", fontsize=12)  
         plt.show()
```



Заключение

- ▶ С развитием технологий даже в простых вещах нужно иметь базовые знания программирования, чтобы успевать за прогрессом. Именно поэтому изучение искусственного интеллекта на языке Python является актуальной и приоритетной темой в сфере информационных технологий и программирования.
- ▶ В изучении алгоритмов машинного обучения и искусственного интеллекта можно поставить долгосрочную цель на проектирование подобия свёрточных нейронных сетей.