

Лабораторная работа №2 по курсу "Технологии машинного обучения"

Выполнила Попова Дарья, студентка группы РТ5-61Б

Масштабирование данных

MinMaxScaler

```
import numpy as np
import pandas as pd
heart = pd.read_csv('C:\\Users\\Дасуиц\\Downloads\\heart.csv')
```

```
heart.head()
```

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

```
from sklearn.preprocessing import MinMaxScaler, StandardScaler, Normalizer
```

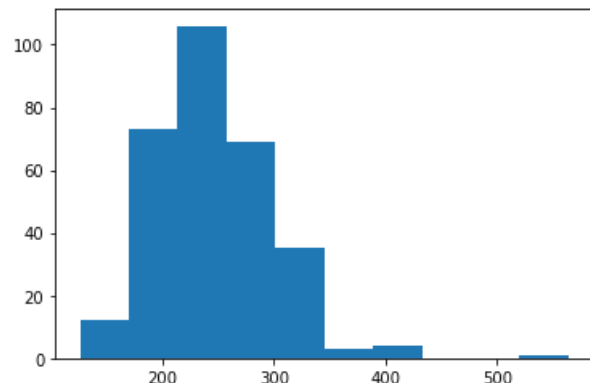
Нормализуем колонку с уровнем холестерина с помощью MinMax масштабирования.

```
minmax_scaler = MinMaxScaler()
minmax_chol = minmax_scaler.fit_transform(heart[['chol']])
```

```
import matplotlib.pyplot as plt
%matplotlib inline
```

```
plt.hist(heart.chol)
```

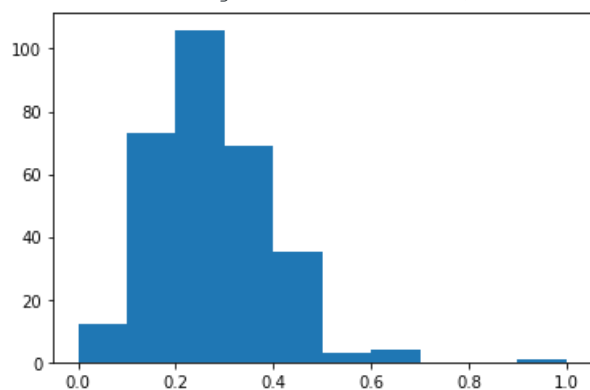
```
(array([ 12.,  73., 106.,  69.,  35.,   3.,   4.,   0.,   0.,   1.]),
array([126. , 169.8, 213.6, 257.4, 301.2, 345. , 388.8, 432.6, 476.4,
       520.2, 564. ]),
<BarContainer object of 10 artists>)
```



```
plt.hist(minmax_chol)
```

Out[15]:

```
(array([ 12.,  73., 106.,  69.,  35.,   3.,   4.,   0.,   0.,   1.]),  
array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),  
<BarContainer object of 10 artists>)
```



StandardScaler

Колонку с показателями верхнего артериального давления в состоянии покоя масштабируем на основе Z-оценки.

In [18]:

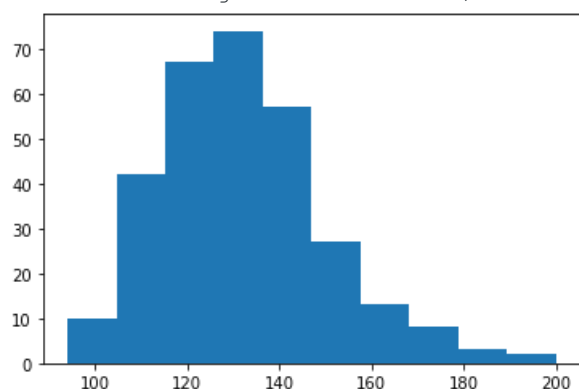
```
standard_scaler = StandardScaler()  
standard_trestbps = standard_scaler.fit_transform(heart[['trestbps']])
```

In [19]:

```
plt.hist(heart.trestbps)
```

Out[19]:

```
(array([10., 42., 67., 74., 57., 27., 13.,  8.,  3.,  2.]),  
array([ 94. , 104.6, 115.2, 125.8, 136.4, 147. , 157.6, 168.2, 178.8,  
       189.4, 200. ]),  
<BarContainer object of 10 artists>)
```

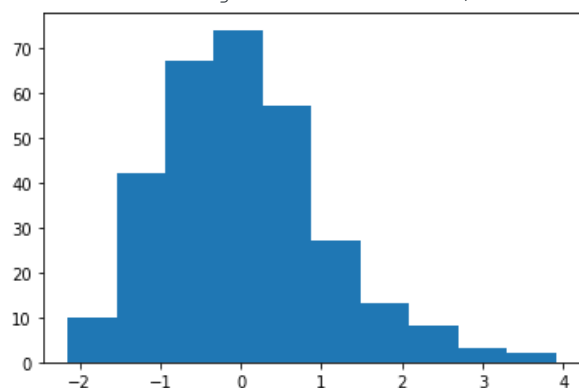


```
plt.hist(standard_trestbps)
```

In [20]:

Out[20]:

```
(array([10., 42., 67., 74., 57., 27., 13.,  8.,  3.,  2.]),  
array([-2.14880232, -1.54340554, -0.93800876, -0.33261198,  0.2727848 ,  
       0.87818158,  1.48357836,  2.08897514,  2.69437191,  3.29976869,  
       3.90516547]),  
<BarContainer object of 10 artists>)
```



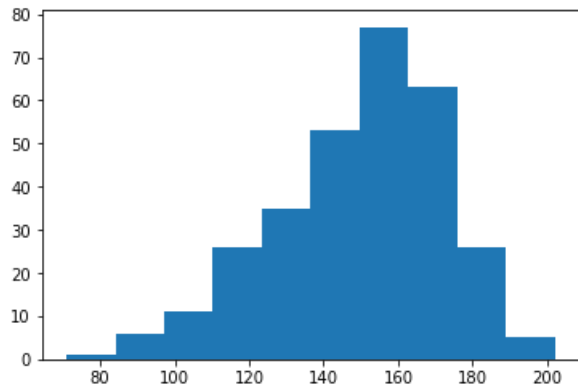
Нормализация

Попробуем также нормализовать колонку с максимальным достигнутым числом сокращений сердечной мышцы в минуту.

```
normalizer = Normalizer()
```

```
plt.hist(heart.thalach)
```

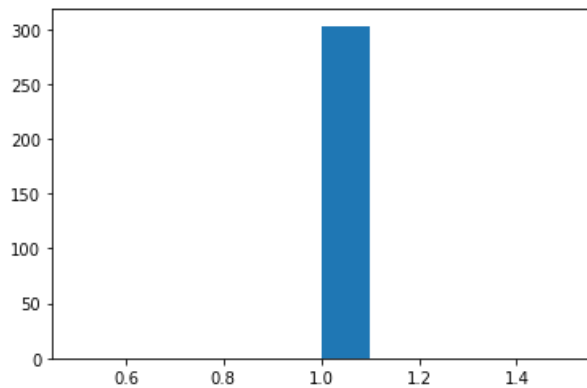
```
(array([ 1.,  6., 11., 26., 35., 53., 77., 63., 26.,  5.]),  
array([ 71. ,  84.1,  97.2, 110.3, 123.4, 136.5, 149.6, 162.7, 175.8,  
       188.9, 202. ]),  
<BarContainer object of 10 artists>)
```



```
norm_data = normalizer.fit_transform(heart[['thalach']])
```

```
plt.hist(norm_data)
```

```
(array([ 0.,  0.,  0.,  0.,  0., 303.,  0.,  0.,  0.,  0.]),  
array([0.5, 0.6, 0.7, 0.8, 0.9, 1. , 1.1, 1.2, 1.3, 1.4, 1.5]),  
<BarContainer object of 10 artists>)
```



Результат, конечно, поражает воображение...

In [21]:

In [29]:

Out[29]:

In [33]:

In [34]:

Out[34]: