Лаборатораня работа №4: Линейные модели, SVM и деревья решений.

```
In [21]:
               #Датасет содержит данные о кредитах на покупку электроники, которые были
              import pandas as pd
              import numpy as np
              from matplotlib import pyplot as plt
              import seaborn as sns
              from sklearn.model selection import train test split, GridSearchCV, Rando
              from sklearn.neighbors import KNeighborsClassifier
              from sklearn.preprocessing import MinMaxScaler, StandardScaler
              from sklearn.linear model import LogisticRegression, LogisticRegressionC
              from sklearn.metrics import accuracy score, precision score, recall score
              from sklearn.tree import DecisionTreeClassifier, export graphviz, export
              from sklearn.svm import SVC
              import graphviz
              from IPython.display import Image
              from warnings import simplefilter
              simplefilter('ignore')
 In [2]:
               # записываем CSV-файл в объект DataFrame
              data = pd.read csv('credit train preprocess.csv', encoding='cp1251', separate pd.read csv('credit train preprocess.csv')
 In [3]:
               # смотрим на первые пять строк
              data.head()
                  age credit_sum credit_month tariff_id score_shk monthly_income
                                                                                                  credit_count
 Out[3]:
             0 34.0
                          59998.00
                                                           1.6
                                                                  0.461599
                                                                                        30000.0
                                                 10
                                                                                                            1.0
             1 34.0
                          10889.00
                                                           1.1
                                                                  0.461599
                                                                                        35000.0
                                                                                                            2.0
             2 32.0
                          10728.00
                                                 12
                                                                  0.461599
                                                                                        35000.0
                                                                                                            5.0
                                                           1.1
             3 27.0
                          12009.09
                                                                  0.461599
                                                                                        35000.0
                                                           1.1
                                                                                                            2.0
               45.0
                          21229.00
                                                 10
                                                                  0.421385
                                                                                        35000.0
                                                                                                            1.0
```

1) Корреляционный анализ

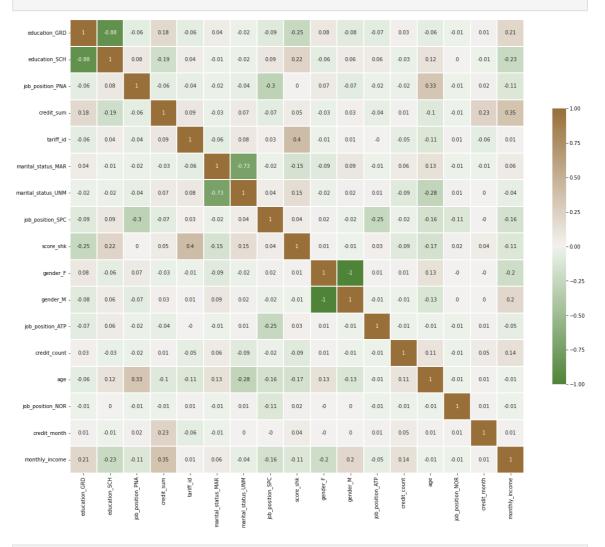
5 rows × 39 columns

```
In [4]:
    corr = data.corr().round(2)
    f, ax = plt.subplots(figsize=(20, 20))
    cmap = sns.diverging_palette(120, 50, as_cmap=True)
    sns.heatmap(data=corr, cmap=cmap, annot=True, vmax=1.0, square=True, line
    plt.show()
```

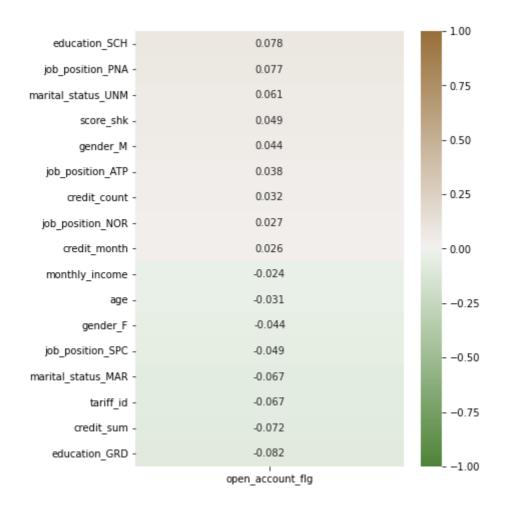
```
credit sum -0.1 1 023 009 00 5 0.35 001 0.01 0.07 0.03 0.04 0.07 0 0.08 0.01 0.01 0 0.01 0 0.00 0.01 0 0.07 0.03 0.03 0.04 0.07 0 0.08 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.01 0 0.
                                                          credit month -0.01 0.23 1 -0.060.04 0.01 0.05 0.01 0.03 -0 0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0
                                                                   tariff_id -0.110.09-0.06 1 0.4 0.010.050.08-0.070.010.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.03-0.02 0 0.01 -0 0.01-0.0 0.04 0.04 0.02 0.020.060.08-0.01
                                                                 score_shk <-0.170.05 0.04 0.4 1 0.110.090.16 0.05 0.01 0.010.03 0.01 0 0.01 0 0.02 0 0 0 0 0 0.010.04 0.04 0 0.02 0 0.010.25 0.04 022 0.06 0.01 0.15 0.15 0.15 0.01
                                                      monthly_income -0.010.35 0.01 0.01-0.11 1 0.14 0.04 0.02 0.2 0.2 0.05 0.14 0 0.16 0 0.04 0 0.01 0 0.11 0 0.01 0 0.01 0 0.01 0.00 1.60 0.13 0.01 0.0 0 0.03 0.21 0.05 0.23 0.02 0.01 0.01 0.00 0.04 0.05
                                                           1.00
                                                   gender_F -0.13-0.03 -0.01001 -0.2 -0.01 -0.004 1 -1 -0.01-0.030.01-0.040.01-0.01 -0.0 -0.07-0.010.01-0.010.02-0.02 -0.02 -0.010.08-0.010.060.020.02 -0.11-0.090.020.12
                                                                gender_M -0.130.03 0 0.01-0.01 02 -0.01 0 0.04 1 1 -0.010.030.010.04-0.010.01 -0 0 0 0.070.01-0.010.02.02 0 0.02 0 0.01-0.080.01.006.002-0.02-0.12
                                                                                                                                                                                                                                                                                                                                                                                                                                   - 0.75
                                                     job_position_ATP -0.010.040.01 -0 0.034.050.01 -0 0.04.001.001 1 -0.02 -0 -0.02 -0 -0 -0.01 -0 -0.02 -0 -0 -0.250.040.01.01 -0 -0 -0.07 -0 0.060.01 -0 0.014.010.01.001
                                                     - 0.50
                                                     job_position_DIR -0.05 0.08-0.01 -0 -0.040.16 0.03 0.01-0.01-0.040.04-0.02-0.03 -0 11 -0 -0.01 -0 -0.01 -0 -0.02 -0 -0 -0.29 0.05 0.01-0.01 -0 -0.02 -0 1 -0.03 -0.1 -0 -0 0.010.03 0.03 0.01
                                                    job_position_INP-0010.010.01 -0 -0.010.04.0.010.01.0.01.0.01.0.01.0.01 -0 -0.01 -0 -0.01 -0 -0.01 -0 -0.01 -0 -0.01 -0 -0.01 -0 -0.01 -0 -0.01 -0 -0.01 -0 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -
                                                                                                                                                                                                                                                                                                                                                                                                                                   -0.25
                                                     job_position_NOR -0.010.010.010.010.02.0.01.01 0 0.03 0 0 0.010.01 0 0.01 0 0 0 1 0 0 0 0 0.110.02 0 0 0 0 0.010.01 0 0.010.01 0 0.010.01 0
                                                    job_position_PNA - 0.33-0.060.02-0.04 0 - 0.11-0.02-0.01-0.08-0.07-0.07-0.02-0.03 - 0.02-0 - 0.01 - 0.01 - 0.01 - 0.01 - 0.01 - 0.01 - 0.03-0.05-0.01-0.01 - 0.0-0.06 - 0.08-0.03-0.01-0.01-0.02-0.04-0.17
                                                     pb position_SPC -0.160.07 -0 0.030.040.160.020.010.050.020.020.020.250.360.050.290.010.070.010.11 -0 -0.3-0.040.020.03 1 0.64-0.090.120.050.020.090.040.090.01.0010.010.020.040.05
                                                    job_position_UMN - -0 007-0 010.020.040.13 0.03 0.01 0.01-0.020.02-0.040.060.01-0.05 -0 -0.01 -0 -0.02 -0 -0.05-0.01 -0 -0.01 -0.05 -1 -0.020.02-0.010.01 013 0.03-0.13-0.01-0.01 -0 0.01-0.01-0.02
                                                                                                                                                                                                                                                                                                                                                                                                                                    -0.50
                                                    job_position_WRK -0.010.01 0 0.010.02-0.010.01.001 0 -0.020.02-0.010.01 0 -0.01 0 -0.01 0 -0.01 0 -0 -0 -0 -0 -0 -0.01 0 -0 -0.01 0 -0 -0.01 0 -0 -0.01 0 -0 -0.01 0 -0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 0 -0.01 
                                                   education_GRD -0.06018 0.01-0.060.25021 0.03 0.02-0.080.08 0.08 0.07 0.02 0 0 1 0.01 0 0 0.060.01 0 0 0.090.13 0.02 0.03 0.01 0.02 1 0.05 0.8 0.21 0.02 0 0.04 0.02 0.05
                                                        education_UGR <-0.13002-0.010.040.06002-0.020.010.01-0.020.020.01-0.01001 0 0.01001 0 0.01 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.010 1 0 0.0
                                                  marital status CIV -0.02 -0 0.010.020.01-0.010.01 0 0.010.02-0.02 -0 0 0.01 -0 0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.01 -0 0.02 -0 -0 -0.020.010.010.01 1 -0.050.180.11-0.02
                                                 marital_status_DIV-011.0.03 -0 -0.02001-0.01003 001001 011-0.11001 0 -0 -0.01 -0 -0 -0 -0 -0 -0 -0.01001 0 -0 -0.01 0 -0 -0 0 0 0 -0 -0 0.01-0.020.05 1 -0.370.220.05
                                                marital_status_MAR -0.13-0.030.010.060.150.06.0.06.002.0.070.090.09-0.010.03 -0 0.030.01 -0 -0.01 -0.0.02 -0 -0.001-0.020.010.010 -0 -0.04 -0.0.04 -0.0.04 -0.010.060.180.37 1 072-0.16
                                               marital_status_UNM -0.280.07 0 0.080.15-0.040.090.030.060.020.02 0.01-0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -0.03 0 -
                                                 acredit sum - credit sum - credit month - monthly income - credit count - credit 
  In [5]:
                                                 print('Признаки, имеющие максимальную по модулю корреляцию с целевым при
                                                 best params = data.corr()['open_account_flg'].map(abs).sort_values(ascen-
                                                 best_params = best_params[best_params.values > 0.02]
                                                 best params
                                            Признаки, имеющие максимальную по модулю корреляцию с целевым признаком
Out[5]: education_GRD 0.082371
                                            education SCH
                                                                                                                                                         0.078337
                                                                                                                                                        0.076889
                                             job_position_PNA
                                                                                                                                                        0.072039
                                             credit_sum
                                             tariff id
                                                                                                                                                             0.067346
                                             marital_status_MAR 0.067112
                                             marital_status_UNM 0.061312
                                             job_position_SPC 0.049143
                                             score_shk
                                                                                                                                                           0.048686
                                             gender F
                                                                                                                                                           0.044265
                                             gender_M
                                                                                                                                                              0.044265
                                             job_position_ATP 0.038288
                                             credit_count
                                                                                                                                                        0.032374
                                                                                                                                                          0.031062
                                             age
                                             job position_NOR
                                                                                                                                                         0.027320
                                                                                                                                                          0.025809
                                             credit month
                                             monthly income
                                                                                                                                                                0.023697
                                             Name: open account flg, dtype: float64
 In [6]:
                                                corr = data[best params.index].corr().round(2)
                                                 f, ax = plt.subplots(figsize=(20, 20))
                                                  cmap = sns.diverging palette(120, 50, as cmap=True)
```

age - 1 0.1 0010.110.170.010.11 0.02-0.030.13-0.130.010.04 0 0.05 0 0.010.01-0.01 0 0.33 0.010.02.00.2-0.16 0 0 0.010.05 0 0.06 0 0.12-0.130.020.11 0.13-0.280.23

sns.heatmap(data=corr, cmap=cmap, annot=True, vmax=1.0, square=True, line
plt.show()



In [7]: plt.figure(figsize=(6, 8))
 sns.heatmap(pd.DataFrame(data[np.append(best_params.index.values, 'open_
 plt.show()



2) Разделение выборки на обучающую и тестовую

```
In [8]:
    data_best = data[best_params.index]
    data_best.head()
```

Out[8]:		education_GRD	education_SCH	job_position_PNA	credit_sum	tariff_id	marital_status_M
	0	1	0	0	59998.00	1.6	
	1	0	1	0	10889.00	1.1	
	2	0	1	0	10728.00	1.1	
	3	0	1	0	12009.09	1.1	
	4	0	1	0	21229.00	1.1	

```
In [9]: y = data['open_account_flg']
   #X = data.drop('open_account_flg', axis=1)
   X = data_best
   x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.75
   x_train, x_test, y_train, y_test = train_test_split(x_train, y_train, test_split(x_train, test_split(x_tr
```

3) Масштабирование данных

```
In [10]:
    scaler = MinMaxScaler().fit(x_train)
    x_train = pd.DataFrame(scaler.transform(x_train), columns=x_train.column
```

```
x_test = pd.DataFrame(scaler.transform(x_test), columns=x_train.columns)
x_train.describe()
```

Out[10]:		education_GRD	education_SCH	job_position_PNA	credit_sum	tariff_id	mari
	count	29880.000000	29880.000000	29880.000000	29880.000000	29880.000000	
	mean	0.425000	0.514759	0.023561	0.117340	0.345539	
	std	0.494351	0.499790	0.151679	0.082275	0.252486	
	min	0.000000	0.000000	0.000000	0.000000	0.000000	
	25%	0.000000	0.000000	0.000000	0.060249	0.106383	
	50%	0.000000	1.000000	0.000000	0.092536	0.340426	
	75%	1.000000	1.000000	0.000000	0.148270	0.638298	
	max	1.000000	1.000000	1.000000	1.000000	1.000000	

4) Логистическая регрессия

```
In [11]:
         # обучаем логистическую регрессию
          clf = LogisticRegression()
          clf.fit(x_train, y_train);
In [12]:
          #считаем точность (долю правильных ответов) на тестовой выборке
          clf.score(x_test, y_test)
         0.8191472747149774
Out[12]:
In [13]:
          def print_metrics(y_test, y_pred):
              print(f"Precision: {precision_score(y_test, y_pred)}")
              print(f"F1-measure: {f1 score(y test, y pred)}")
In [14]:
          # вычисляем метрики точности
          y_pred_log = clf.predict(x_test)
          print metrics(y test, y pred log)
         Precision: 0.5862068965517241
         F1-measure: 0.01446808510638298
```

5) SVC

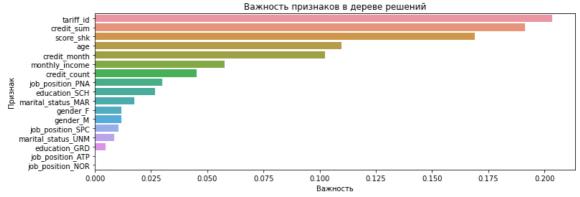
```
In [15]:
    best_svm_model = SVC()
    best_svm_model.fit(x_train, y_train)
    y_pred_svm = best_svm_model.predict(x_test)
    print_metrics(y_test, y_pred_svm)
```

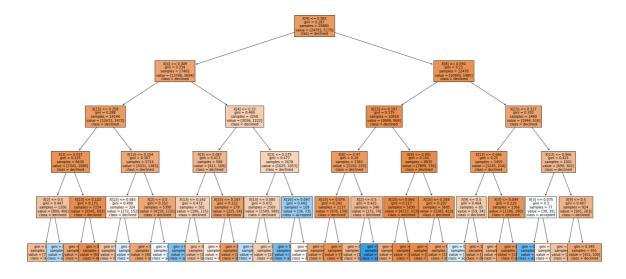
Precision: 0.5714285714285714 F1-measure: 0.003436426116838488

6) Дерево решений

```
In [16]:
```

```
params = {'min samples leaf': range(3, 30)}
          tree = DecisionTreeClassifier(random state=3) #max depth=5
          grid cv = GridSearchCV(estimator=tree, cv=5, param grid=params, n jobs=-
          grid cv.fit(x train, y train)
           #tree.fit(x train, y train)
          print(grid cv.best params )
          {'min samples leaf': 29}
In [17]:
          best tree = grid cv.best estimator
          best tree.fit(x train, y train)
          y pred tree = best tree.predict(x test)
          print metrics(y test, y pred tree)
          Precision: 0.4374255065554231
          F1-measure: 0.23227848101265824
In [18]:
          importances = pd.DataFrame(data=zip(x_train.columns, best_tree.feature_in
          print('Важность признаков в дереве решений\n')
          for row in importances.sort values(by='Важность', ascending=False).value
              print(f'{row[0]}: {round(row[1], 3)}')
          Важность признаков в дереве решений
          tariff id: 0.203
          credit sum: 0.191
          score_shk: 0.169
          age: 0.11
          credit month: 0.102
          monthly_income: 0.058
          credit count: 0.045
          job position PNA: 0.03
          education_SCH: 0.027
          marital status MAR: 0.017
          gender F: 0.012
          gender_M: 0.012
          job position SPC: 0.01
          marital status UNM: 0.009
          education GRD: 0.005
          job position ATP: 0.0
          job position NOR: 0.0
In [19]:
          plt.figure(figsize=(12, 4))
          sns.barplot(data=importances.sort values(by='Важность', ascending=False)
          plt.title('Важность признаков в дереве решений')
          plt.show()
                                          Важность признаков в дереве решений
                 tariff_id
               credit sum
                score_shk
                   age
             credit_month
monthly income
```





7) Сравнение моделей

```
In [32]:
          print('Логистическая регрессия')
          print_metrics(y_test, y_pred_log)
          print('\nMeтoд опорных векторов')
          print_metrics(y_test, y_pred_svm)
          print('\nДерево решений')
          print_metrics(y_test, y_pred_tree)
         Логистическая регрессия
         Precision: 0.5862068965517241
         F1-measure: 0.01446808510638298
         Метод опорных векторов
         Precision: 0.5714285714285714
         F1-measure: 0.003436426116838488
         Дерево решений
         Precision: 0.4374255065554231
         F1-measure: 0.23227848101265824
 In [ ]:
```