Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Системы обработки информации и управления

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К КУРСОВОЙ РАБОТЕ

НА ТЕМУ:

Решение комплексной задачи машинного обучения с учителем

Студент __ИУ-5 33М__ _____ __**А.П.Корнеева**__
(Группа) (Подпись, дата) (И.О.Фамилия)

Преподователь _____ __**Ю.Е.Гапанюк**
(Подпись, дата) (И.О.Фамилия)

2021 г.

Содержание

Введение

Аннотация

В ходе выполнения курсовой работы был проведён анализ выбранного набора данных, а также произведено предсказание размеров суммы трат покупателя. В работе содержится 17 страниц и 27 рисунков.

Актуальность

Результаты исследования являются востребованными и актуальными, т.к. компаниям, продающим свою продукцию, необходимо иметь результаты прогноза относительно размеров предполагаемой суммы трат покупателя, чтобы рассчитывать стоимость товаров.

Цель работы

Решить комплексную задачу машинного обучения с учителем.

Методы

Анализ и синтез.

Основная часть

В ходе работы было произведено исследование набора данных, чтобы удалить имеющиеся пропуски для исключения искажения прогноза. Далее было выполнено кодирование категориальных признаков, а именно:

- «Gender»;
- «Partner»;
- «Dependents»;
- «PhoneService»;
- «MultipleLines»;
- «InternetService»;
- «OnlineSecurity»;
- «OnlineBackup»;
- «DeviceProtection»;
- «TechSupport»;
- «StreamingTV»;
- «StreamingMovies»;
- «Contract»;
- «PaperlessBilling»;
- «PaymentMethod»;
- «Churn».

Следующим шагом являлось обучение модели с различными вариантами. Было произведено построение графиков метрик качества модели, а также модели с использованием AutoML.

Предсказание суммы трат пользователя, в зависимости от его интересов

In [1]:
```
from google.colab import drive
drive.mount('/content/drive', force_remount=True)
path = "/content/drive/My Drive/Colab Notebooks/ммо/dz"
```

```
Mounted at /content/drive
```

In [2]:
```
!pip3 install tensorflow==2.2
```

```
Requirement already satisfied: tensorflow==2.2 in /usr/local/lib/python3.7/dist-packages (2.2.0)
Requirement already satisfied: gast==0.3.3 in /usr/local/lib/python3.7/dist-packages (from tensorflow==2.2) (0.3.
3)
Requirement already satisfied: keras-preprocessing>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from tensorfl
ow==2.2) (1.1.2)
Requirement already satisfied: numpy<2.0,>=1.16.0 in /root/.local/lib/python3.7/site-packages (from tensorflow==2.
2) (1.20.3)
Requirement already satisfied: protobuf>=3.8.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow==2.2)
(3.12.4)
Requirement already satisfied: absl-py>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow==2.2) (0.
12.0)
Requirement already satisfied: tensorflow-estimator<2.3.0,>=2.2.0 in /usr/local/lib/python3.7/dist-packages (from
tensorflow==2.2) (2.2.0)
Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow==2.2)
(1.1.0)
Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.7/dist-packages (from tensorflow==2.2)
(3.3.0)
Collecting scipy==1.4.1; python_version >= "3"
  Using cached https://files.pythonhosted.org/packages/dd/82/c1fe128f3526b128cfd185580ba40d01371c5d299fcf7f77968e2
2dfcc2e/scipy-1.4.1-cp37-cp37m-manylinux1_x86_64.whl
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow==2.2) (1.15.
0)
Requirement already satisfied: grpcio>=1.8.6 in /usr/local/lib/python3.7/dist-packages (from tensorflow==2.2) (1.3
2.0)
```

```
Installing collected packages: scipy
  Found existing installation: scipy 1.6.1
    Uninstalling scipy-1.6.1:
      Successfully uninstalled scipy-1.6.1
Successfully installed scipy-1.4.1
```

In [3]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.impute import SimpleImputer
from sklearn.impute import MissingIndicator
from sklearn.impute import KNNImputer
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import Lasso
from sklearn.pipeline import Pipeline
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestRegressor
from sklearn.experimental import enable_iterative_imputer
from sklearn.impute import IterativeImputer
from IPython.display import Image
import scipy.stats as stats
%matplotlib inline
sns.set(style="ticks")
```

In [4]:

```
data = pd.read_csv(path+'/telecom_users.csv')
```

In [5]:

```
data.head()
```

Out[5]:

ection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
ıternet service	No internet service	No internet service	No internet service	Two year	No	Credit card (automatic)	24.10	1734.65	No
Yes	No	Yes	No	Month-to-month	Yes	Credit card (automatic)	88.15	3973.2	No
No	No	No	No	Month-to-month	Yes	Bank transfer (automatic)	74.95	2869.85	Yes
No	No	No	Yes	Month-to-month	Yes	Electronic check	55.90	238.5	No
Yes	No	No	No	Month-to-month	No	Electronic check	53.45	119.5	No

In [6]:

```
data.shape
```

Out[6]:

```
(5986, 22)
```

Пропуски

In [7]:
```
data_features = list(zip(
# признаки
[i for i in data.columns],
zip(
    # типы колонок
    [str(i) for i in data.dtypes],
    # проверим есть ли пропущенные значения
    [i for i in data.isnull().sum()]
)))
# Признаки с типом данных и количеством пропусков
data_features
```

Out[7]:
```
[('Unnamed: 0', ('int64', 0)),
 ('customerID', ('object', 0)),
 ('gender', ('object', 0)),
 ('SeniorCitizen', ('int64', 0)),
 ('Partner', ('object', 0)),
 ('Dependents', ('object', 0)),
 ('tenure', ('int64', 0)),
 ('PhoneService', ('object', 0)),
 ('MultipleLines', ('object', 0)),
 ('InternetService', ('object', 0)),
 ('OnlineSecurity', ('object', 0)),
 ('OnlineBackup', ('object', 0)),
 ('DeviceProtection', ('object', 0)),
 ('TechSupport', ('object', 0)),
 ('StreamingTV', ('object', 0)),
 ('StreamingMovies', ('object', 0)),
 ('Contract', ('object', 0)),
('PaperlessBilling', ('object', 0)),
('PaymentMethod', ('object', 0)),
('MonthlyCharges', ('float64', 0)),
('TotalCharges', ('object', 0)),
('Churn', ('object', 0))]
```

Пропусков в наборе данных не обнаружено.

Кодирование категориальных признаков

In [8]:
```
from sklearn.preprocessing import LabelEncoder
```

In [9]:
```
data_encoded = data.copy()
```

In [10]:
```
columns = ['gender', 'Partner', 'Dependents', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity'
for column in columns:
    label_encoder = LabelEncoder()
    data_encoded[column] = label_encoder.fit_transform(data_encoded[column])
```

In [11]:
```
data_encoded = data_encoded.drop([data_encoded.columns[0], data_encoded.columns[1], data_encoded.columns[-1], data
data_encoded.head()
```

Out[11]:

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	Dev
0	1	0	1	1	72	1	2	2	1	1	
1	0	0	0	0	44	1	0	1	0	2	
2	0	1	1	0	38	1	2	1	0	0	
3	1	0	0	0	4	1	0	0	0	0	
4	1	0	0	0	2	1	0	0	2	0	

In [12]:
```
from sklearn.model_selection import train_test_split
```

In [13]:
```
x = data_encoded.drop([data_encoded.columns[-1]], axis=1)
x
```

Out[13]:

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup
0	1	0	1	1	72	1	2	2	1	1
1	0	0	0	0	44	1	0	1	0	2
2	0	1	1	0	38	1	2	1	0	0
3	1	0	0	0	4	1	0	0	0	0
4	1	0	0	0	2	1	0	0	2	0
...	...	...	...	...	...	...	...	...	...	...
5981	1	0	1	0	1	1	0	1	2	0
5982	0	0	1	1	23	1	2	0	2	2
5983	1	0	1	1	12	1	0	2	1	1
5984	1	1	0	0	12	1	2	1	0	0
5985	1	0	0	0	26	1	0	2	1	1

5986 rows × 17 columns

In [14]:
```
def arr_to_df(arr_scaled):
    res = pd.DataFrame(arr_scaled, columns=x.columns)
    return res
```

In [15]:
```
# Разделим выборку на обучающую и тестовую
X_train, X_test, y_train, y_test = train_test_split(x, data_encoded['MonthlyCharges'],
                                                    test_size=0.2,
                                                    random_state=1)
# Преобразуем массивы в DataFrame
X_train_df = arr_to_df(X_train)
X_test_df = arr_to_df(X_test)

X_train_df.shape, X_test_df.shape, y_train.shape, y_test.shape
```

Out[15]:
```
((4788, 17), (1198, 17), (4788,), (1198,))
```

In [16]:
```
class MetricLogger:

    def __init__(self):
        self.df = pd.DataFrame(
            {'metric': pd.Series([], dtype='str'),
            'alg': pd.Series([], dtype='str'),
            'value': pd.Series([], dtype='float')})

    def add(self, metric, alg, value):
        """
        Добавление значения
        """
        # Удаление значения если оно уже было ранее добавлено
        self.df.drop(self.df[(self.df['metric']==metric)&(self.df['alg']==alg)].index, inplace = True)
        # Добавление нового значения
        temp = [{'metric':metric, 'alg':alg, 'value':value}]
        self.df = self.df.append(temp, ignore_index=True)

    def get_data_for_metric(self, metric, ascending=True):
        """
        Формирование данных с фильтром по метрике
        """
        temp_data = self.df[self.df['metric']==metric]
        temp_data_2 = temp_data.sort_values(by='value', ascending=ascending)
        return temp_data_2['alg'].values, temp_data_2['value'].values
```

```
    def plot(self, str_header, metric, ascending=True, figsize=(5, 5)):
        """
        Вывод графика
        """
        array_labels, array_metric = self.get_data_for_metric(metric, ascending)
        fig, ax1 = plt.subplots(figsize=figsize)
        pos = np.arange(len(array_metric))
        rects = ax1.barh(pos, array_metric,
                         align='center',
                         height=0.5,
                         tick_label=array_labels)
        ax1.set_title(str_header)
        for a,b in zip(pos, array_metric):
            plt.text(0.5, a-0.05, str(round(b,3)), color='white')
        plt.show()
```

In [17]:
```
from sklearn.linear_model import LinearRegression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error
```

In [18]:
```
clas_models_dict = {'LinR': LinearRegression(),
                    'SVR': SVR(),
                    'KNN_5':KNeighborsRegressor(n_neighbors=5),
                    'Tree':DecisionTreeRegressor(random_state=1),
                    'GB': GradientBoostingRegressor(random_state=1),
                    'RF':RandomForestRegressor(n_estimators=50, random_state=1)}
```

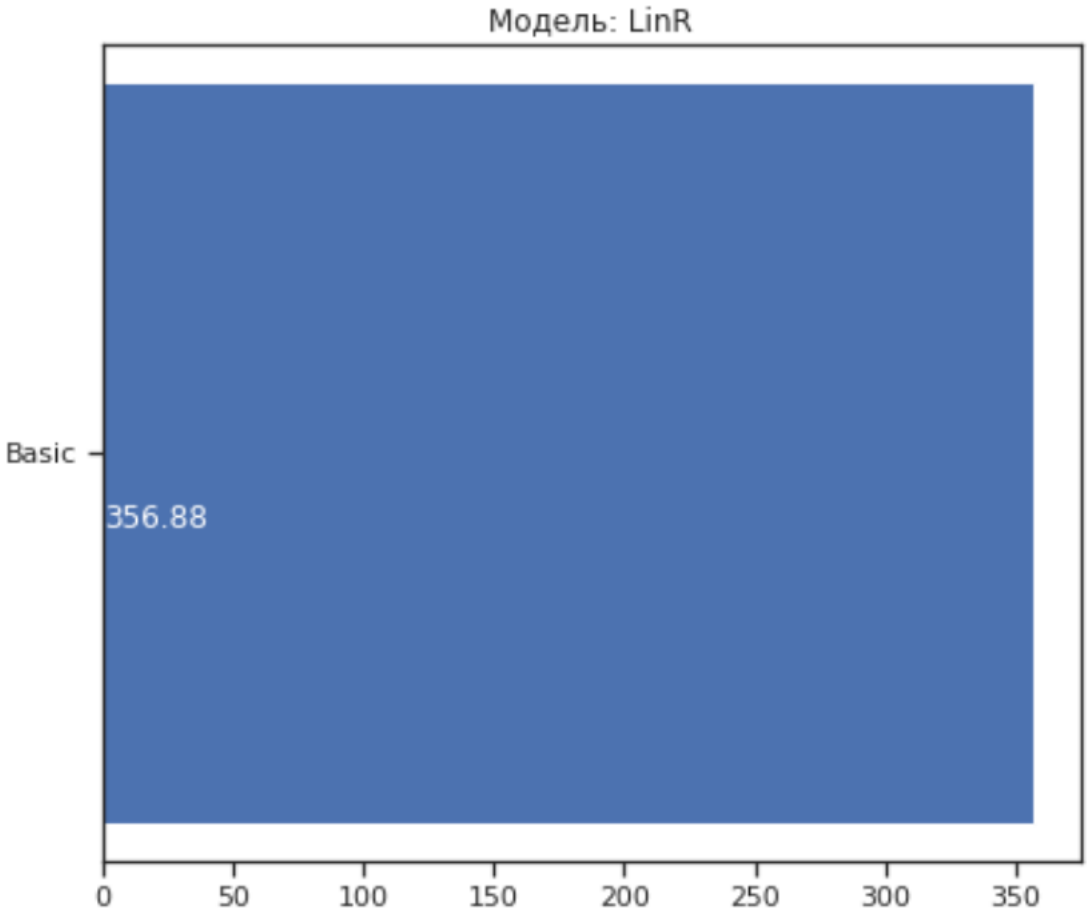
In [19]:
```
X_data_dict = {'Basic': (X_train_df, X_test_df)}
```

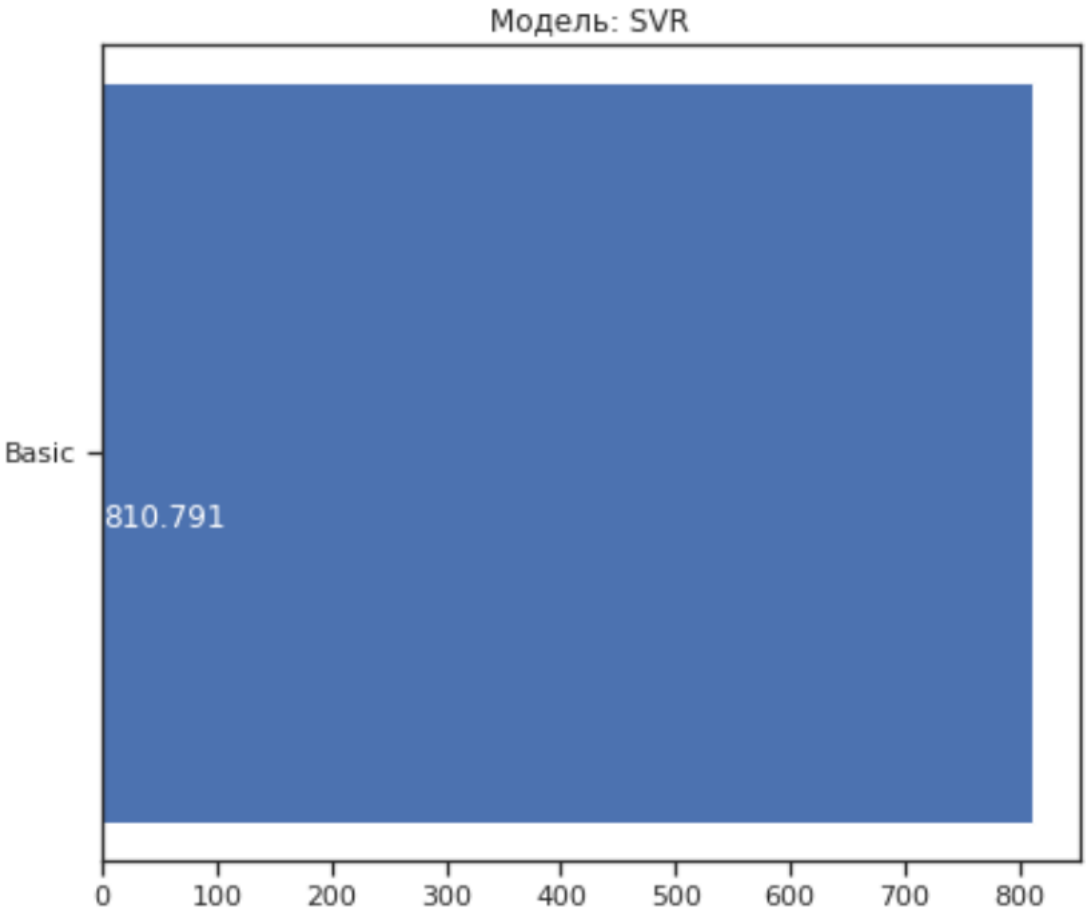
In [20]:
```
def test_models(clas_models_dict, X_train, X_test, y_train, y_test):

    logger = MetricLogger()

    for model_name, model in clas_models_dict.items():
        model.fit(X_train, y_train)
        y_pred = model.predict(X_test)
        mse = mean_squared_error(y_test, y_pred)
        logger.add(model_name, 'Basic', mse)

    return logger
```

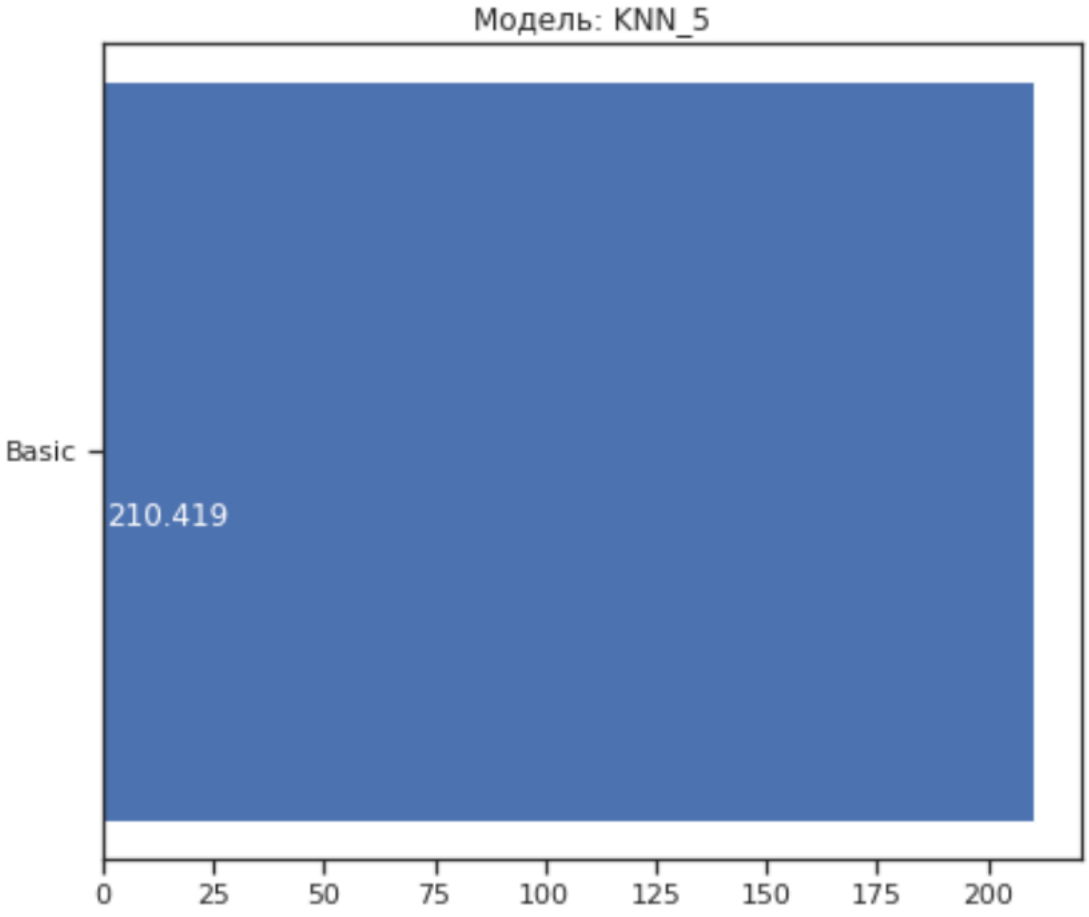
In [21]:
```
logger = test_models(clas_models_dict, X_train_df, X_test_df, y_train, y_test)
```

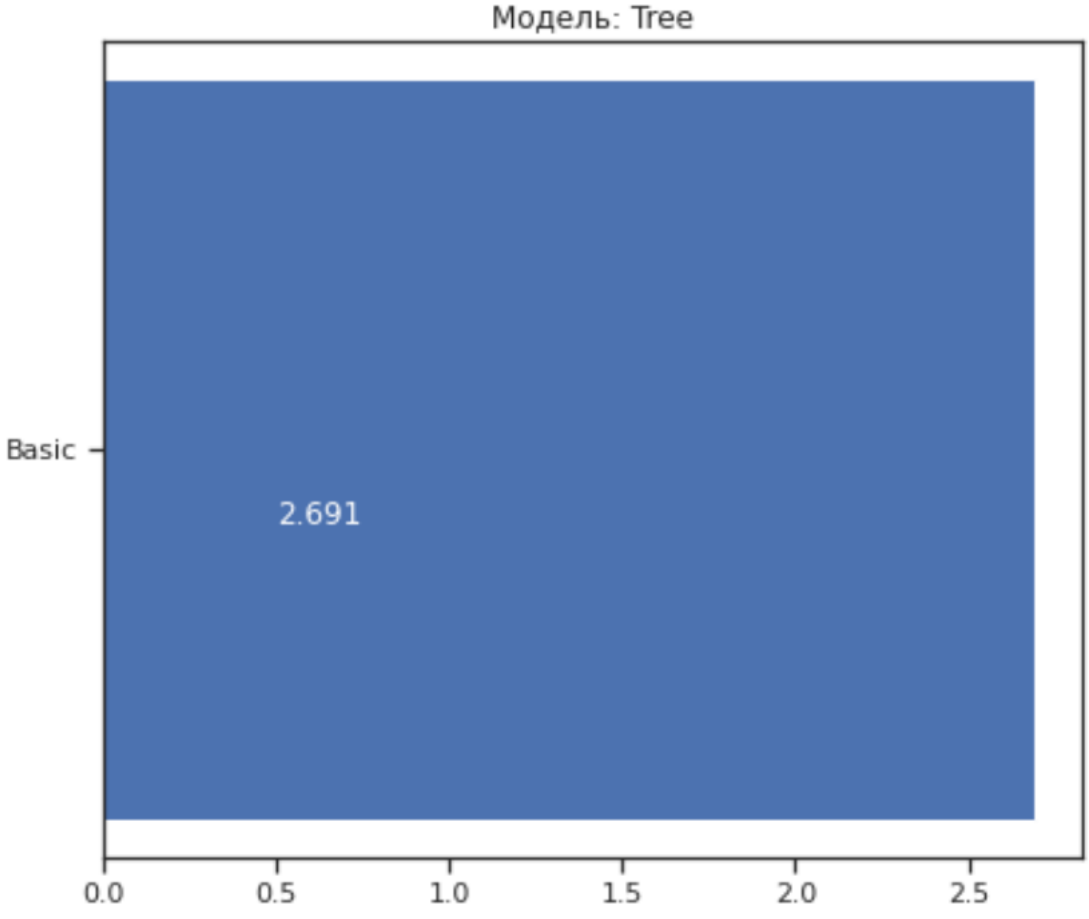
Построение графика метрик качества модели

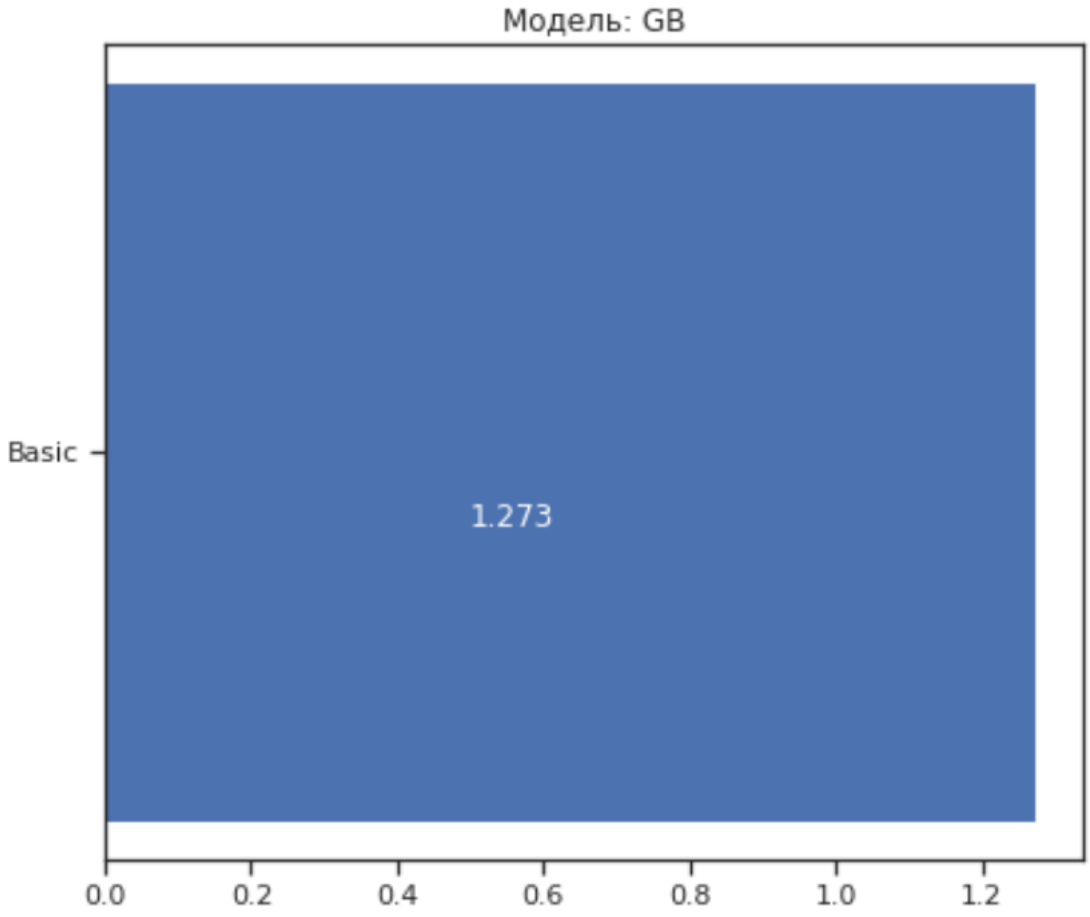
In [22]:
```
# Построим графики метрик качества модели
for model in clas_models_dict:
    logger.plot('Модель: ' + model, model, figsize=(7, 6))
```

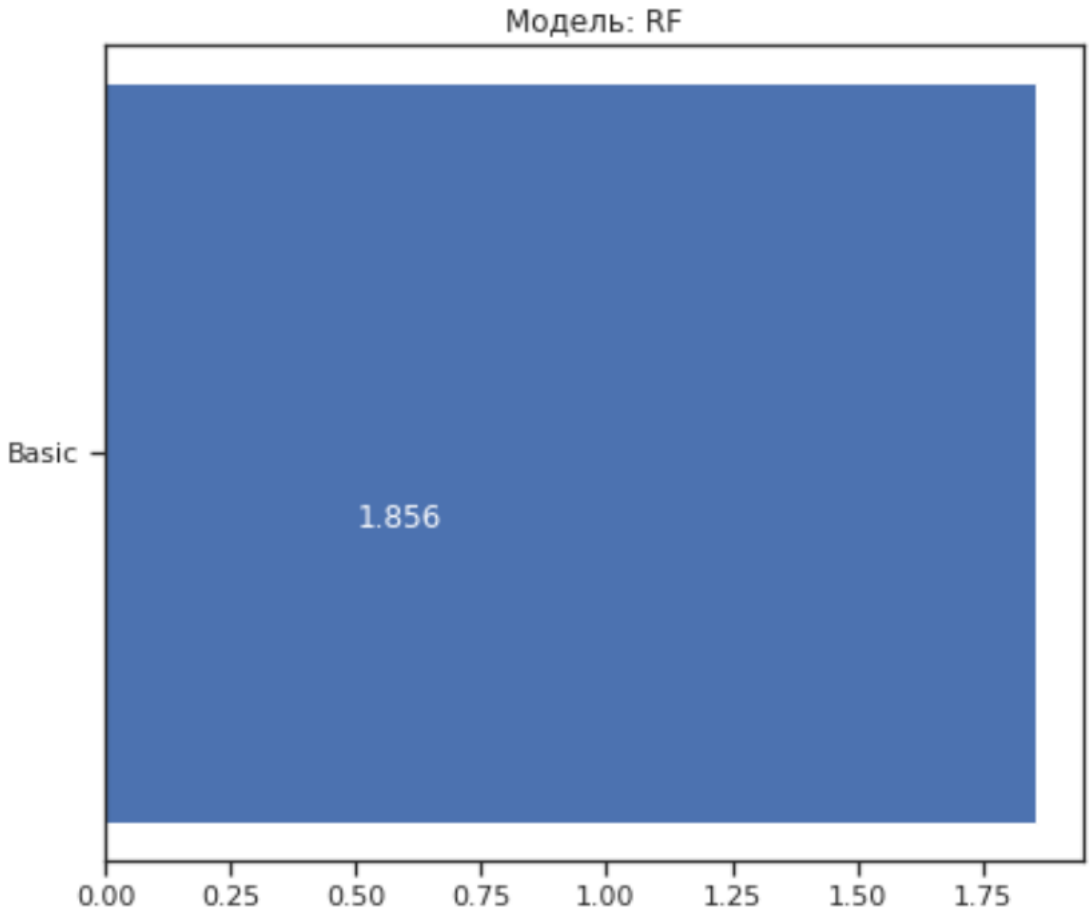



Модель: SVR
Basic
810.791
0
100
200
300
400
500
600
700
800


Модель: KNN_5
Basic
210.419
0
25
50
75
100
125
150
175
200


Модель: Tree
Basic
2.691
0.0
0.5
1.0
1.5
2.0
2.5


Модель: GB
Basic
1.273
0.0
0.2
0.4
0.6
0.8
1.0
1.2



AutoML

In [25]:

```
!pip3 install --user mljar-supervised
!pip3 install delayed
```

```
Requirement already satisfied: mljar-supervised in /root/.local/lib/python3.7/site-packages (0.10.4)
Requirement already satisfied: category-encoders==2.2.2 in /usr/local/lib/python3.7/dist-packages (from mljar-supe
rvised) (2.2.2)
Requirement already satisfied: wordcloud==1.8.1 in /root/.local/lib/python3.7/site-packages (from mljar-supervise
d) (1.8.1)
Requirement already satisfied: tabulate==0.8.7 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(0.8.7)
Requirement already satisfied: cloudpickle==1.3.0 in /usr/local/lib/python3.7/dist-packages (from mljar-supervise
d) (1.3.0)
Requirement already satisfied: scikit-plot==0.3.7 in /root/.local/lib/python3.7/site-packages (from mljar-supervis
ed) (0.3.7)
Requirement already satisfied: dtreeviz==1.3 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(1.3)
Requirement already satisfied: scikit-learn==0.24.2 in /root/.local/lib/python3.7/site-packages (from mljar-superv
ised) (0.24.2)
Requirement already satisfied: catboost==0.24.4 in /root/.local/lib/python3.7/site-packages (from mljar-supervise
d) (0.24.4)
Requirement already satisfied: pandas==1.2.0 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(1.2.0)
Requirement already satisfied: markdown in /usr/local/lib/python3.7/dist-packages (from mljar-supervised) (3.3.4)
Requirement already satisfied: xgboost==1.3.3 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(1.3.3)
```

```
Requirement already satisfied: lightgbm==3.0.0 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(3.0.0)
Requirement already satisfied: optuna==2.7.0 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(2.7.0)
Requirement already satisfied: seaborn==0.11.1 in /usr/local/lib/python3.7/dist-packages (from mljar-supervised)
(0.11.1)
Requirement already satisfied: matplotlib>=3.2.2 in /usr/local/lib/python3.7/dist-packages (from mljar-supervised)
(3.2.2)
Requirement already satisfied: joblib==1.0.1 in /usr/local/lib/python3.7/dist-packages (from mljar-supervised) (1.
0.1)
Requirement already satisfied: pyarrow>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from mljar-supervised)
(3.0.0)
Requirement already satisfied: scipy==1.6.1 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(1.6.1)
Requirement already satisfied: numpy>=1.20.0 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(1.20.3)
Requirement already satisfied: shap==0.36.0 in /root/.local/lib/python3.7/site-packages (from mljar-supervised)
(0.36.0)
Requirement already satisfied: patsy>=0.5.1 in /usr/local/lib/python3.7/dist-packages (from category-encoders==2.
2.2->mljar-supervised) (0.5.1)
Requirement already satisfied: statsmodels>=0.9.0 in /usr/local/lib/python3.7/dist-packages (from category-encoder
s==2.2.2->mljar-supervised) (0.10.2)
Requirement already satisfied: pillow in /usr/local/lib/python3.7/dist-packages (from wordcloud==1.8.1->mljar-supe
rvised) (7.1.2)
Requirement already satisfied: pytest in /usr/local/lib/python3.7/dist-packages (from dtreeviz==1.3->mljar-supervi
sed) (3.6.4)
Requirement already satisfied: colour in /root/.local/lib/python3.7/site-packages (from dtreeviz==1.3->mljar-super
vised) (0.1.5)
Requirement already satisfied: graphviz>=0.9 in /usr/local/lib/python3.7/dist-packages (from dtreeviz==1.3->mljar-
supervised) (0.10.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in /root/.local/lib/python3.7/site-packages (from scikit-learn
==0.24.2->mljar-supervised) (2.1.0)
Requirement already satisfied: plotly in /usr/local/lib/python3.7/dist-packages (from catboost==0.24.4->mljar-supe
rvised) (4.4.1)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from catboost==0.24.4->mljar-supervi
sed) (1.15.0)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas==1.2.
0->mljar-supervised) (2.8.1)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (from pandas==1.2.0->mljar-s
upervised) (2018.9)
Requirement already satisfied: importlib-metadata; python_version < "3.8" in /usr/local/lib/python3.7/dist-package
s (from markdown->mljar-supervised) (4.0.1)
Requirement already satisfied: cmaes>=0.8.2 in /root/.local/lib/python3.7/site-packages (from optuna==2.7.0->mljar
-supervised) (0.8.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from optuna==2.7.0->mlja
r-supervised) (20.9)
Requirement already satisfied: alembic in /root/.local/lib/python3.7/site-packages (from optuna==2.7.0->mljar-supe
rvised) (1.6.2)
Requirement already satisfied: colorlog in /root/.local/lib/python3.7/site-packages (from optuna==2.7.0->mljar-sup
ervised) (5.0.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from optuna==2.7.0->mljar-supervise
d) (4.41.1)
Requirement already satisfied: sqlalchemy>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from optuna==2.7.0->ml
jar-supervised) (1.4.15)
Requirement already satisfied: cliff in /root/.local/lib/python3.7/site-packages (from optuna==2.7.0->mljar-supe
rvised) (3.7.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=3.2.2->m
ljar-supervised) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=3.
2.2->mljar-supervised) (1.3.1)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-package
s (from matplotlib>=3.2.2->mljar-supervised) (2.4.7)
Requirement already satisfied: slicer in /root/.local/lib/python3.7/site-packages (from shap==0.36.0->mljar-supe
rvised) (0.0.7)
Requirement already satisfied: numba in /usr/local/lib/python3.7/dist-packages (from shap==0.36.0->mljar-supervi
sed) (0.51.2)
Requirement already satisfied: atomicwrites>=1.0 in /usr/local/lib/python3.7/dist-packages (from pytest->dtreevi
z==1.3->mljar-supervised) (1.4.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from pytest->dtreeviz==1.3-
>mljar-supervised) (56.1.0)
Requirement already satisfied: more-itertools>=4.0.0 in /usr/local/lib/python3.7/dist-packages (from pytest->dtr
eeviz==1.3->mljar-supervised) (8.7.0)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.7/dist-packages (from pytest->dtreeviz==
1.3->mljar-supervised) (21.2.0)
Requirement already satisfied: py>=1.5.0 in /usr/local/lib/python3.7/dist-packages (from pytest->dtreeviz==1.3->
mljar-supervised) (1.10.0)
Requirement already satisfied: pluggy<0.8,>=0.5 in /usr/local/lib/python3.7/dist-packages (from pytest->dtreeviz
==1.3->mljar-supervised) (0.7.1)
Requirement already satisfied: retrying>=1.3.3 in /usr/local/lib/python3.7/dist-packages (from plotly->catboost=
=0.24.4->mljar-supervised) (1.3.3)
Requirement already satisfied: typing-extensions>=3.6.4; python_version < "3.8" in /usr/local/lib/python3.7/dist
-packages (from importlib-metadata; python_version < "3.8"->markdown->mljar-supervised) (3.7.4.3)
```

```
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata; pyt
hon_version < "3.8"->markdown->mljar-supervised) (3.4.1)
Requirement already satisfied: Mako in /root/.local/lib/python3.7/site-packages (from alembic->optuna==2.7.0->ml
jar-supervised) (1.1.4)
Requirement already satisfied: python-editor>=0.3 in /root/.local/lib/python3.7/site-packages (from alembic->opt
una==2.7.0->mljar-supervised) (1.0.4)
Requirement already satisfied: greenlet!=0.4.17; python_version >= "3" in /usr/local/lib/python3.7/dist-packages
(from sqlalchemy>=1.1.0->optuna==2.7.0->mljar-supervised) (1.1.0)
Requirement already satisfied: cmd2>=1.0.0 in /root/.local/lib/python3.7/site-packages (from cliff->optuna==2.7.
0->mljar-supervised) (1.5.0)
Requirement already satisfied: PrettyTable>=0.7.2 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna=
=2.7.0->mljar-supervised) (2.1.0)
Requirement already satisfied: pbr!=2.1.0,>=2.0.0 in /root/.local/lib/python3.7/site-packages (from cliff->optun
a==2.7.0->mljar-supervised) (5.6.0)
Requirement already satisfied: PyYAML>=3.12 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna==2.7.0
->mljar-supervised) (3.13)
Requirement already satisfied: stevedore>=2.0.1 in /root/.local/lib/python3.7/site-packages (from cliff->optuna=
=2.7.0->mljar-supervised) (3.3.0)
Requirement already satisfied: llvmlite<0.35,>=0.34.0.dev0 in /usr/local/lib/python3.7/dist-packages (from numba
->shap==0.36.0->mljar-supervised) (0.34.0)
Requirement already satisfied: MarkupSafe>=0.9.2 in /usr/local/lib/python3.7/dist-packages (from Mako->alembic->
optuna==2.7.0->mljar-supervised) (2.0.0)
Requirement already satisfied: wcwidth>=0.1.7 in /usr/local/lib/python3.7/dist-packages (from cmd2>=1.0.0->cliff
->optuna==2.7.0->mljar-supervised) (0.2.5)
Requirement already satisfied: colorama>=0.3.7 in /root/.local/lib/python3.7/site-packages (from cmd2>=1.0.0->cl
iff->optuna==2.7.0->mljar-supervised) (0.4.4)
Requirement already satisfied: pyperclip>=1.6 in /root/.local/lib/python3.7/site-packages (from cmd2>=1.0.0->cli

ff->optuna==2.7.0->mljar-supervised) (1.8.2)
Requirement already satisfied: delayed in /usr/local/lib/python3.7/dist-packages (0.11.0b1)
Requirement already satisfied: hiredis in /usr/local/lib/python3.7/dist-packages (from delayed) (2.0.0)
Requirement already satisfied: redis in /usr/local/lib/python3.7/dist-packages (from delayed) (3.5.3)
```

In [26]:
```
from supervised.automl import AutoML
```

In [32]:
```
train = data
train.head()
```

Out[32]:

	Unnamed: 0	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineS
0	1869	7010-BRBUU	Male	0	Yes	Yes	72	Yes	Yes	No	No i
1	4528	9688-YGXVR	Female	0	No	No	44	Yes	No	Fiber optic	
2	6344	9286-DOJGF	Female	1	Yes	No	38	Yes	Yes	Fiber optic	
3	6739	6994-KERXL	Male	0	No	No	4	Yes	No	DSL	
4	432	2181-UAESM	Male	0	No	No	2	Yes	No	DSL	

In [28]:
```
automl = AutoML()
```

In [36]:
```
automl.fit(train[train.columns[2:-3]], train['MonthlyCharges'])
```

```
AutoML directory: AutoML_1
The task is regression with evaluation metric rmse
AutoML will use algorithms: ['Baseline', 'Linear', 'Decision Tree', 'Random Forest', 'Xgboost', 'Neural Networ
k']
AutoML will ensemble availabe models
AutoML steps: ['simple_algorithms', 'default_algorithms', 'ensemble']
* Step simple_algorithms will try to check up to 3 models
1_Baseline rmse 30.041484 trained in 0.42 seconds
2_DecisionTree rmse 9.566765 trained in 13.15 seconds
3_Linear rmse 36.208657 trained in 3.66 seconds
* Step default_algorithms will try to check up to 3 models
4_Default_Xgboost rmse 1.103866 trained in 7.6 seconds
5_Default_NeuralNetwork rmse 2.882078 trained in 1.61 seconds
6_Default_RandomForest rmse 6.037609 trained in 9.99 seconds
* Step ensemble will try to check up to 1 model
Ensemble rmse 1.103866 trained in 0.37 seconds
/root/.local/lib/python3.7/site-packages/numpy/lib/function_base.py:2642: RuntimeWarning:

invalid value encountered in true_divide

/root/.local/lib/python3.7/site-packages/numpy/lib/function_base.py:2643: RuntimeWarning:

invalid value encountered in true_divide

AutoML fit time: 49.05 seconds
AutoML best model: 4_Default_Xgboost
```

Вывод

В ходе выполнения курсовой работы была проведена проверка набора данных на предмет наличия пропусков, было произведено кодирование категориальных признаков, а также был построен график метрик качества модели и произведена работа с AutoML. Следовательно, цель работы достигнута.

Список литературы

1. https://www.automl.org/
2. Python: категориальные признаки | Анализ малых данных (dyakonov.org)
3. Метрики в задачах машинного обучения / Хабр (habr.com)