



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Прогнозирование доходов крестьянских (фермерских) хозяйств

Дашиева Баярма Шагдаровна



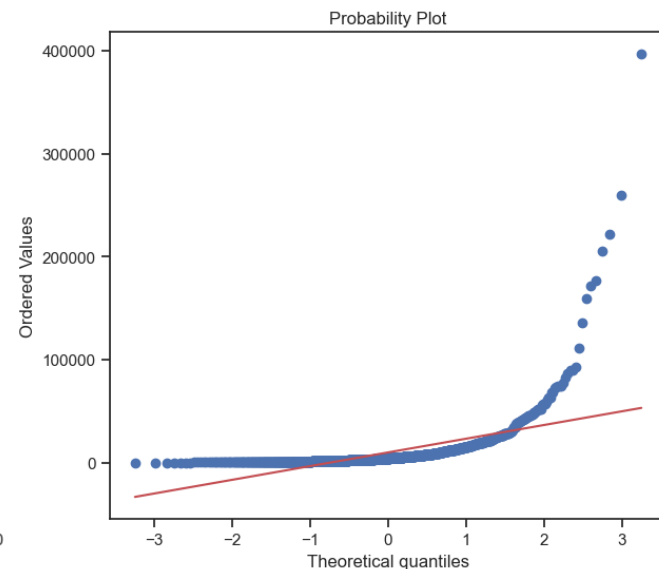
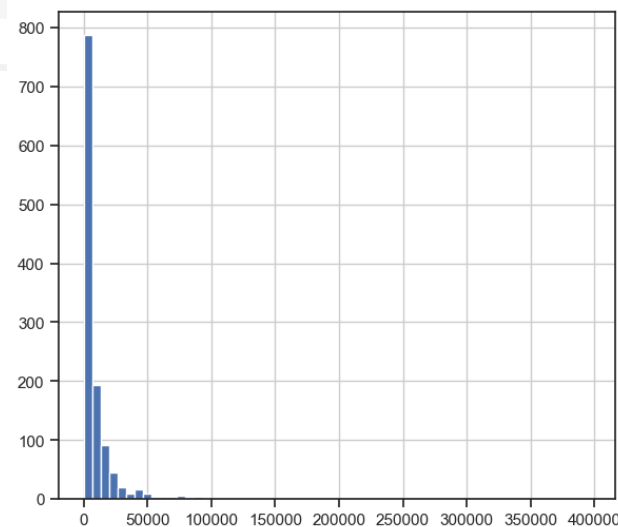
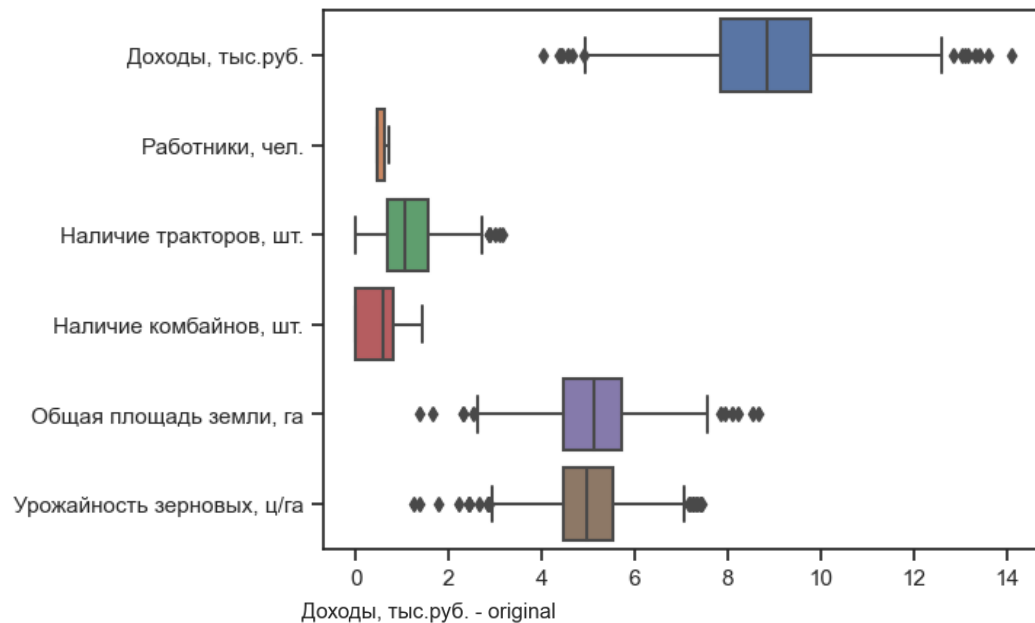
Данные по крестьянским (фермерским) хозяйствам

Out[4]:

№		Доходы, руб - за 2019 год, руб.	в том числе: от реализации сельскохозяйственной продукции, продуктов ее первичной и промышленной переработки - за 2019 год	от оказания услуг - за 2019 год	получено средств государственной поддержки (субсидии, гранты) - за 2019 год	Расходы, тыс. руб - за 2019 год	в том числе:	из них: техника, машины и оборудование - за 2019 год	племенные и продуктивные животные - за 2019 год	земельные участки - за 2019 год	...	Сельского: техника налич
							расходы на приобретение основных средств, включая лизинговые платежи ln(стр.231211+ 231212+ 231213+ 231214) - за 2019 год					
0	1	111242000.0	84843000	25959000.0	330000.0	73295000	120000.0	120000.0	NaN	NaN	...	
1	2	4734000.0	3572000	NaN	1162000.0	4181000	35000.0	NaN	35000.0	NaN	...	
2	3	6147000.0	6147000	NaN	NaN	4650000	NaN	NaN	NaN	NaN	...	
3	4	4132000.0	4132000	NaN	NaN	4120000	NaN	NaN	NaN	NaN	...	
4	5	2676000.0	2676000	NaN	NaN	2112000	228000.0	28000.0	NaN	200000.0	...	

5 rows x 56 columns

Показатель	Количество пропусков
Доходы, руб.	0
Численность постоянных работников, чел	654
Члены КФХ (включая главу КФХ), чел	0
Зерновые и зернобобовые культуры на зерно и семена (кроме рис) - урожайность, ц/га	0
в том числе: тракторы - наличие на начало года	209
в том числе: тракторы - наличие на конец года	186
комбайны - наличие на начало года	507
комбайны - наличие на конец года	494
Земельные участки и объекты природопользования - всего, га - наличие на начало года	10
Земельные участки и объекты природопользования - всего, га - наличие на конец года	0



Гистограмма и график Q-Q распределения КФХ по доходам КФХ



Разведочный анализ данных

In [12]: df_pok.head()

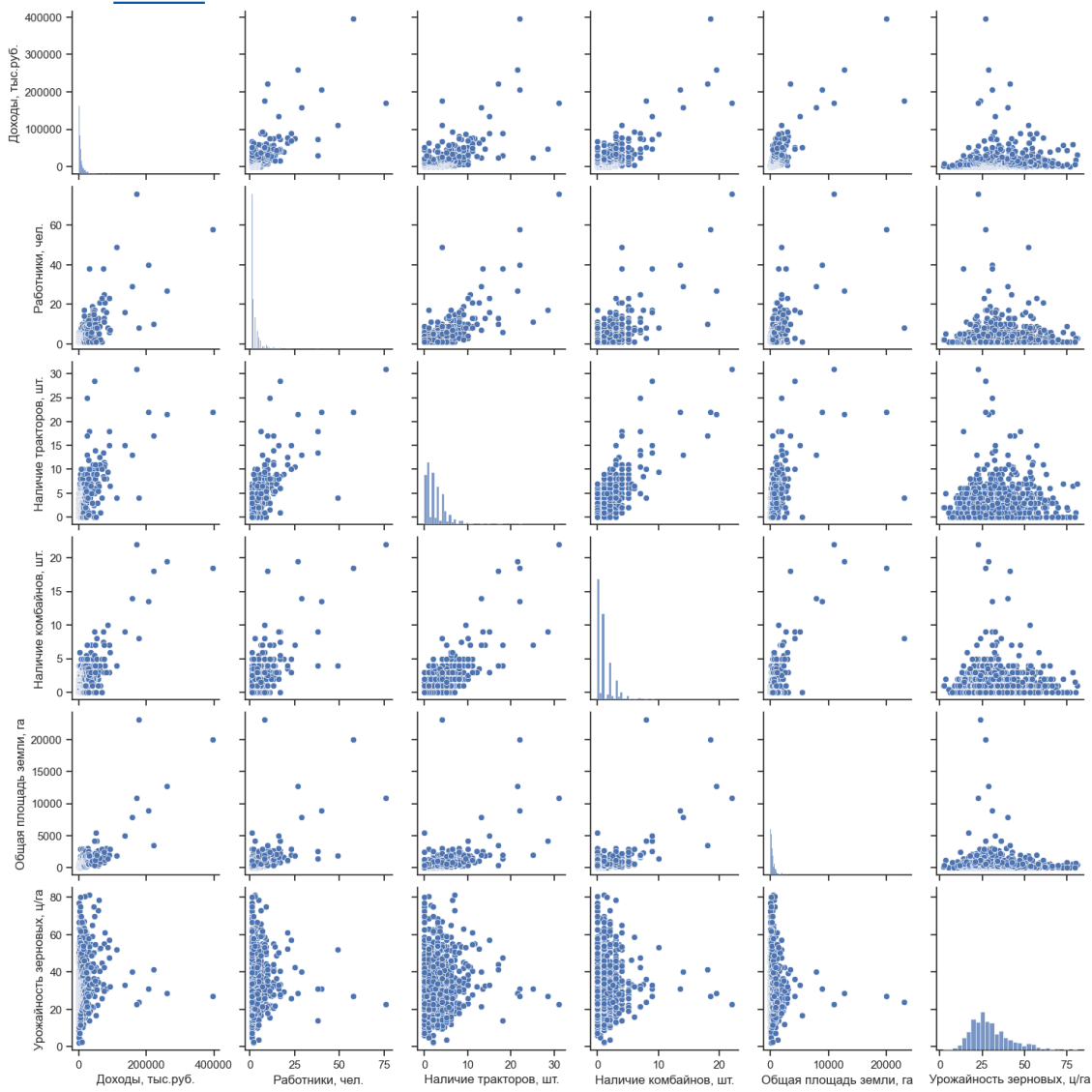
	Доходы, тыс.руб.	Работники, чел.	Наличие тракторов, шт.	Наличие комбайнов, шт.	Общая площадь земли, га	Урожайность зерновых, ц/га
0	111242.0	49.0	4.0	4.0	1854.0	52.2
1	4734.0	1.0	0.0	0.0	762.0	30.3
2	6147.0	2.0	1.0	0.0	450.0	25.2
3	4132.0	4.0	4.0	3.0	861.0	28.8
4	2676.0	3.0	1.0	0.0	120.0	28.6

In [18]: df_pok.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1202 entries, 0 to 1201
Data columns (total 6 columns):
#   Column                                Non-Null Count  Dtype  
---  -
0   Доходы, тыс.руб.                      1202 non-null  float64
1   Работники, чел.                       1202 non-null  float64
2   Наличие тракторов, шт.                 1202 non-null  float64
3   Наличие комбайнов, шт.                 1202 non-null  float64
4   Общая площадь земли, га                1202 non-null  float64
5   Урожайность зерновых, ц/га             1202 non-null  float64
dtypes: float64(6)
memory usage: 56.5 KB
```

In [16]: df_pok.describe()

	Доходы, тыс.руб.	Работники, чел.	Наличие тракторов, шт.	Наличие комбайнов, шт.	Общая площадь земли, га	Урожайность зерновых, ц/га
count	1202.0	1202.0	1202.0	1202.0	1202.0	1202.0
mean	9871.5	2.9	2.8	1.2	504.9	30.4
std	22013.3	4.7	3.0	1.8	1167.7	12.8
min	50.0	1.0	0.0	0.0	3.0	2.0
25%	1717.5	1.0	1.0	0.0	120.0	21.5
50%	4153.0	1.0	2.0	1.0	259.4	27.7
75%	9716.2	3.0	4.0	2.0	539.8	36.4
max	396570.0	76.0	31.0	22.0	23118.5	81.3

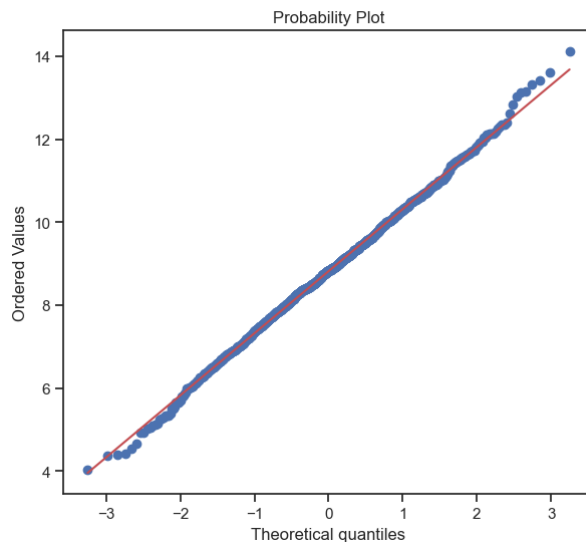
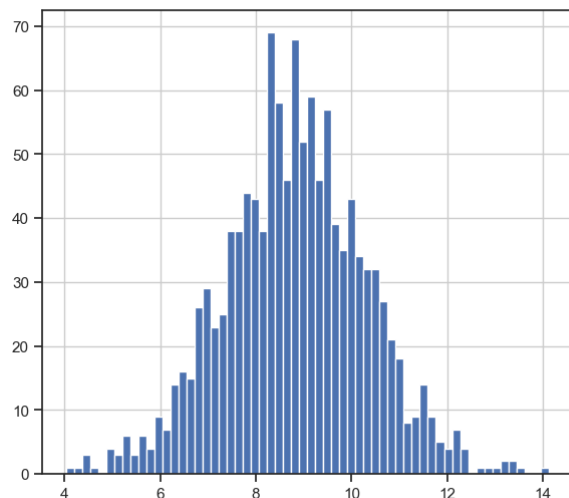




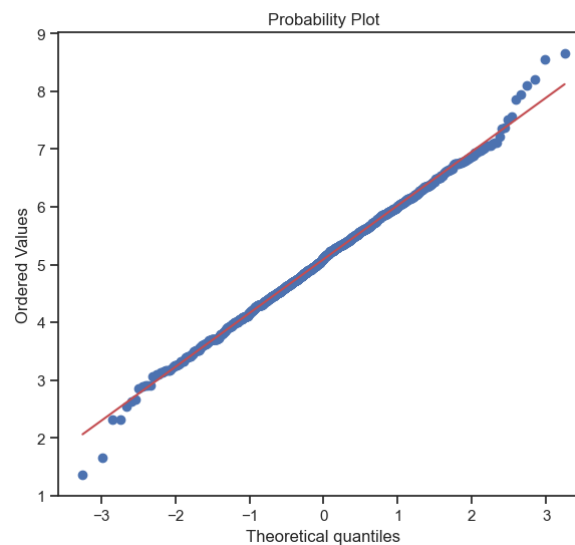
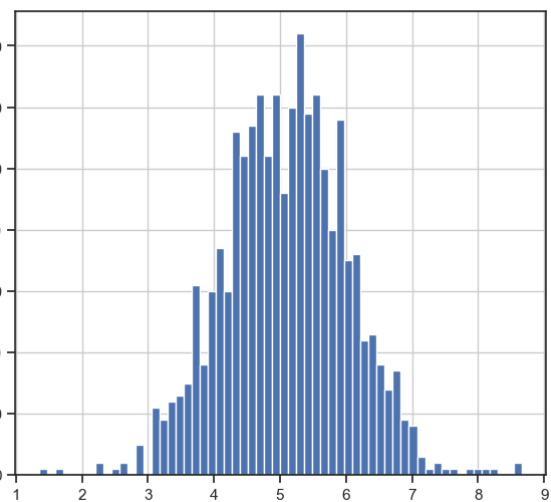
ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГУ им. Н.Э. Баумана

Предобработка данных

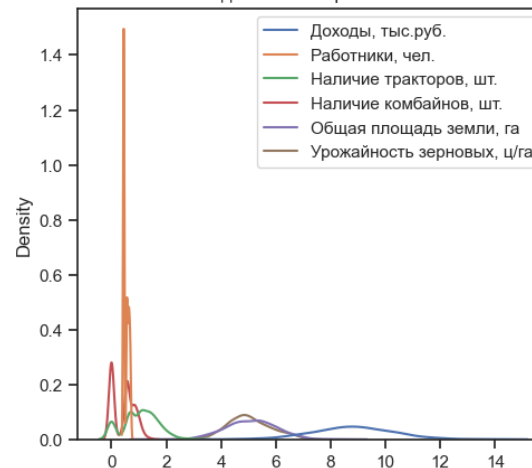
Доходы, тыс.руб. - преобразование Йео-Джонсона



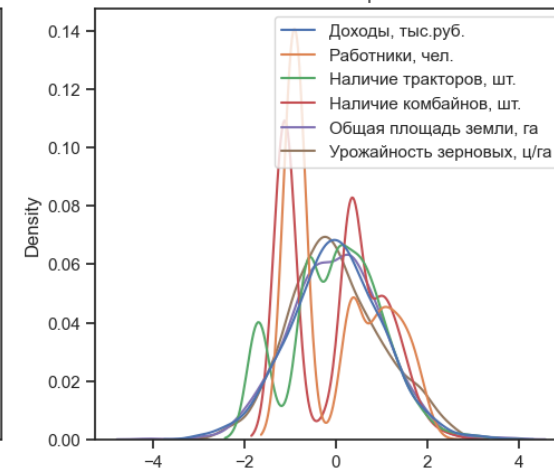
Общая площадь земли, га - преобразование Йео-Джонсона



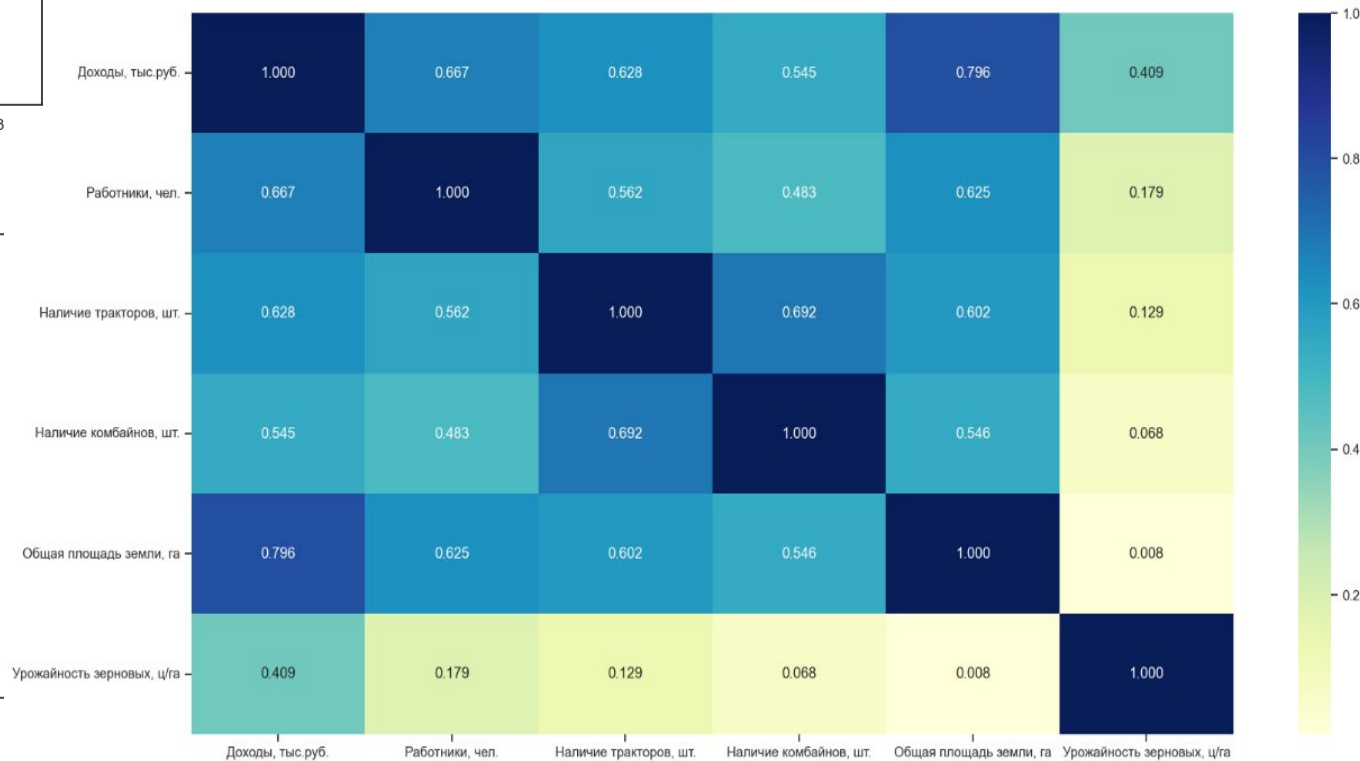
до масштабирования



после масштабирования



Корреляционная матрица Pearson





Out[121]:

OLS Regression Results

Dep. Variable:	Доходы, тыс.руб.	R-squared:	0.707
Model:	OLS	Adj. R-squared:	0.706
Method:	Least Squares	F-statistic:	660.2
Date:	Tue, 25 Apr 2023	Prob (F-statistic):	2.25e-218
Time:	19:06:49	Log-Likelihood:	-671.64
No. Observations:	826	AIC:	1351.
Df Residuals:	822	BIC:	1370.
Df Model:	3		
Covariance Type:	nonrobust		
	coef	std err	t P> t [0.025 0.975]
const	0.0154	0.019	0.807 0.420 -0.022 0.053
Работники, чел.	0.2216	0.026	8.599 0.000 0.171 0.272
Наличие тракторов, шт.	0.1627	0.024	6.661 0.000 0.115 0.211
Общая площадь земли, га	0.5706	0.027	21.203 0.000 0.518 0.623
Omnibus:	11.146	Durbin-Watson:	2.113
Prob(Omnibus):	0.004	Jarque-Bera (JB):	12.636
Skew:	-0.206	Prob(JB):	0.00180
Kurtosis:	3.445	Cond. No.	2.52

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correct.

МЕТОД К-БЛИЖАЙШИХ СОСЕДЕЙ

```
knn = KNeighborsRegressor()
param_grid = {'n_neighbors': [1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20]}
GSCV_knn = GridSearchCV(estimator=knn, param_grid=param_grid, cv=10, verbose=2)
GSCV_knn.fit(X_train, y_train)
GSCV_knn.best_params_
```

ДЕРЕВО РЕШЕНИЙ

```
decision_tree = DecisionTreeRegressor(random_state = 42)
param_grid = {'max_features': ['auto', 'sqrt', 'log2'],
              'max_depth': [3, 4, 5, 6],
              'min_samples_leaf': [1, 2, 3]}
GSCV_dt = GridSearchCV(estimator=decision_tree, param_grid=param_grid, cv=10, verbose=2)
GSCV_dt.fit(X_train, y_train)
GSCV_dt.best_params_
```

СЛУЧАЙНЫЙ ЛЕС

```
In [149]: random_forest = RandomForestRegressor(random_state = 42)
param_grid = {'n_estimators': [100, 200, 300, 400, 500],
              'max_features': ['auto', 'sqrt'],
              'max_depth': [4, 5, 6, 7, 8],
              'criterion': ['squared_error']}
GSCV_rf = GridSearchCV(estimator=random_forest, param_grid=param_grid, cv=10, verbose=2)
GSCV_rf.fit(X_train, y_train)
GSCV_rf.best_params_
```

РИДЖ-РЕГРЕССИЯ

```
In [127]: ridge = Ridge()
param_grid_ridge = [{'alpha': [1e-15, 1e-10, 1e-8, 1e-4, 1e-3, 1e-2, 1e-1, 0, 1, 5, 10],
                    'solver': ['svd', 'cholesky', 'lsqr', 'sparse_cg', 'sag', 'saga']}]
GSCV_ridge = GridSearchCV(estimator=ridge, param_grid=param_grid_ridge, cv=10, verbose=2)
GSCV_ridge.fit(X_train, y_train)
GSCV_ridge.best_params_
```

ЛАССО-РЕГРЕССИЯ

```
In [132]: lasso = Lasso()
param_grid_lasso = {'alpha': [1e-15, 1e-10, 1e-8, 1e-4, 1e-3, 1e-2, 1e-1, 0, 1, 5, 10]}
GSCV_lasso = GridSearchCV(estimator=lasso, param_grid=param_grid_lasso, cv=10, verbose=2)
GSCV_lasso.fit(X_train, y_train)
GSCV_lasso.best_params_
```

	R2_train	R2_test	MSE_train	MSE_test	MAE_train	MAE_test
OLS1	0.707	0.668	0.298	0.318	0.440	0.453
OLS2	0.707	0.668	0.298	0.318	0.440	0.453
Ridge	0.707	0.668	0.298	0.318	0.440	0.453
Lasso	0.707	0.668	0.298	0.318	0.440	0.453
KNN	0.732	0.665	0.272	0.321	0.413	0.449
DT	0.713	0.639	0.291	0.345	0.434	0.474
RF	0.767	0.682	0.237	0.304	0.389	0.444
GB	0.792	0.682	0.211	0.304	0.363	0.440
MLP	0.729	0.683	0.275	0.303	0.417	0.436
NN1	0.695	0.647	0.309	0.338	0.445	0.468
NN2	0.713	0.683	0.291	0.303	0.431	0.440
NN3	0.719	0.685	0.285	0.301	0.426	0.439



Полносвязная нейронная сеть

Архитектура нейронной сети

Model: "sequential_6"

Layer (type)	Output Shape	Param #
dense_18 (Dense)	(None, 90)	360
dropout_3 (Dropout)	(None, 90)	0
dense_19 (Dense)	(None, 80)	7280
dense_20 (Dense)	(None, 1)	81

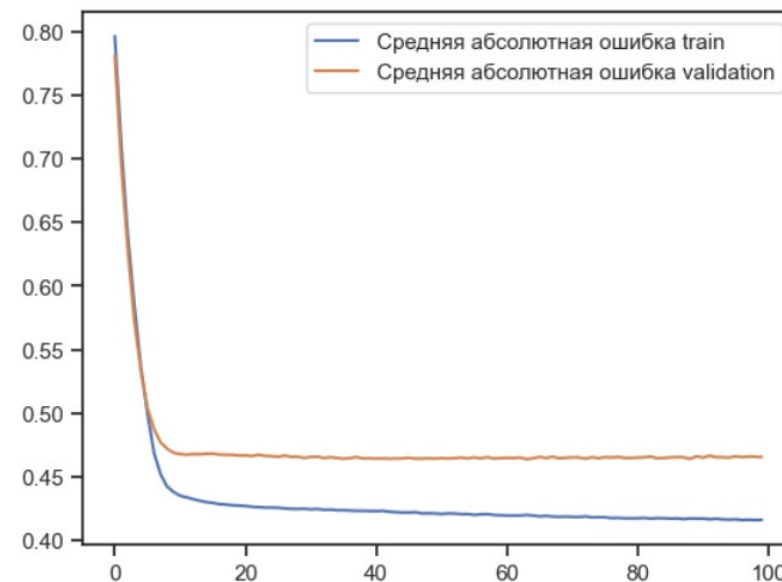
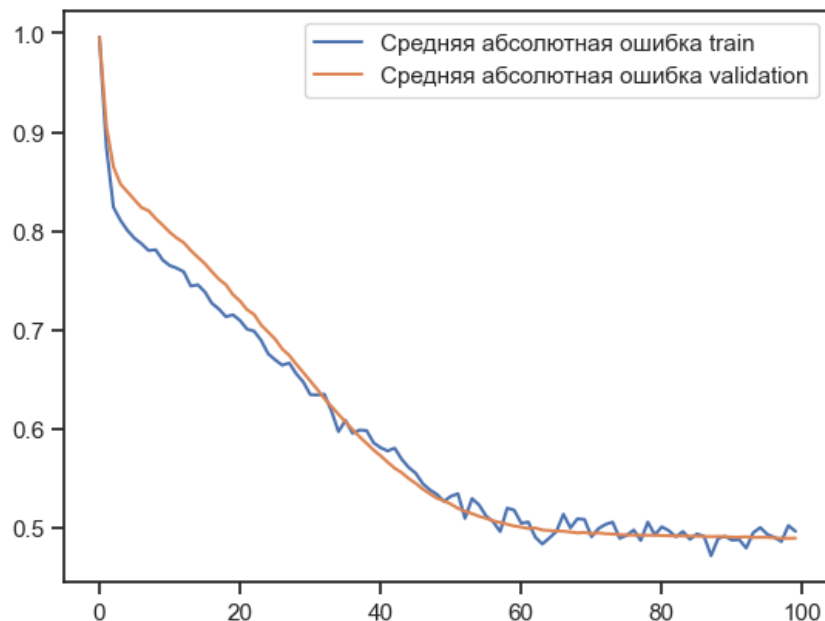
=====
Total params: 7,721
Trainable params: 7,721
Non-trainable params: 0

Model: "sequential_2"

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 90)	360
dense_7 (Dense)	(None, 80)	7280
dense_8 (Dense)	(None, 1)	81

=====
Total params: 7,721
Trainable params: 7,721
Non-trainable params: 0

График функции потерь





Flask-приложение

Для запуска приложения пользователь должен перейти по ссылке на сайт: <http://127.0.0.1:5000/>.

Flask-приложение представляет собой форму, состоящую из трех входов, куда вводятся значения трех параметров: численность работников, чел., число тракторов, шт., площадь земли, га. Введенные значения должны быть больше или равны 0, в противном случае появится ошибка «ОШИБКА! Введенные значения должны быть больше или равны 0».

После этого нужно нажать на кнопку «Submit», и модель выдаст прогнозное значение доходов КФХ при заданных параметрах.

Работники, чел.: 23 Наличие тракторов, шт.: 34 Общая площадь земли, га: 23

Работники, чел.

Наличие тракторов, шт.

Общая площадь земли, га

Submit

ОШИБКА! Введенные значения должны быть больше или равны 0.

Работники, чел.

Наличие тракторов, шт.

Общая площадь земли, га

Submit



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru