Neural networks usage of financial time series prediction

Elena V. Alymova, Oleg E. Kudryavtsev

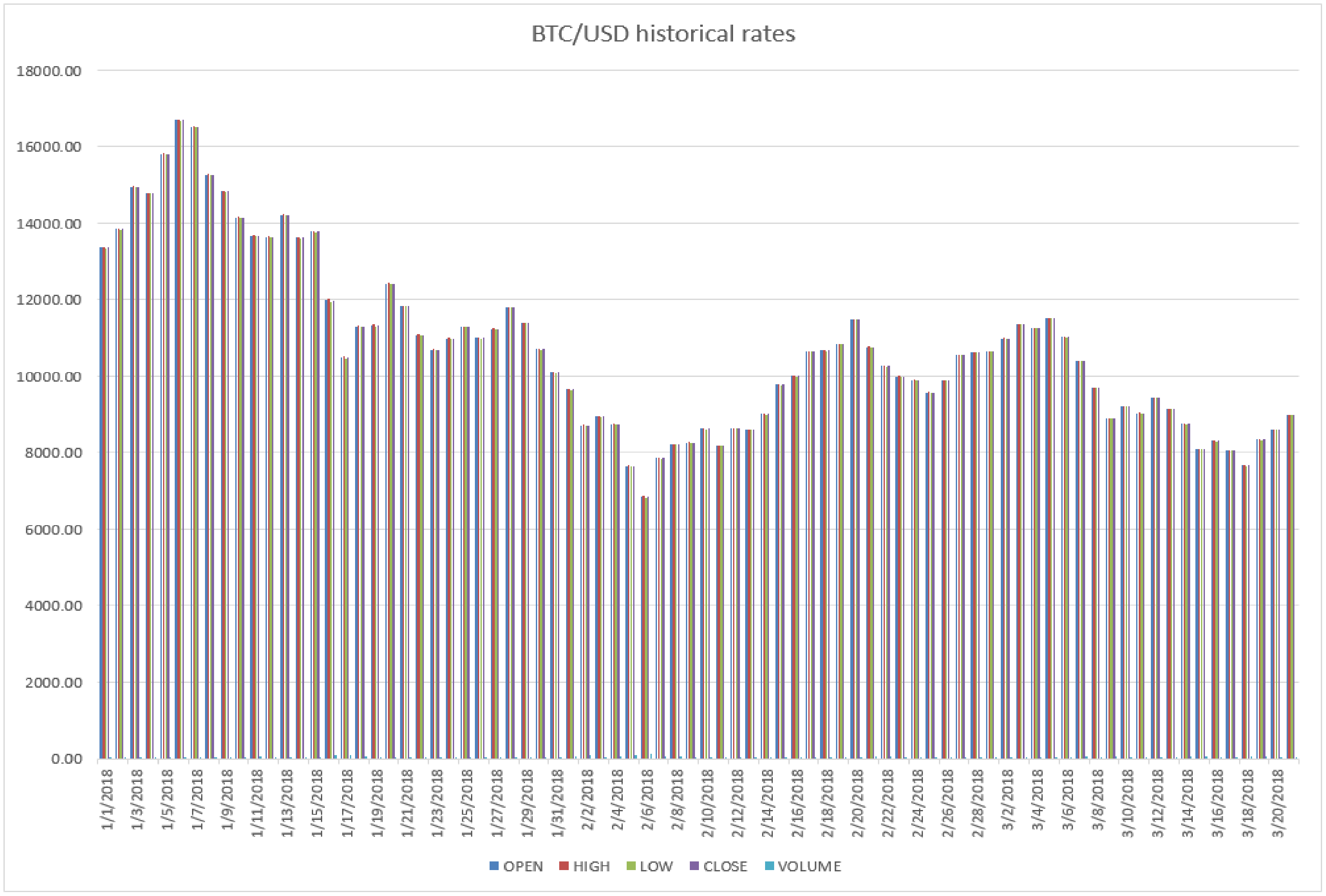
Russian Customs Academy, Rostov-on-Don

4th International Conference on Stochastic Methods
2–9 June 2019, Divnomorskoye , Russia

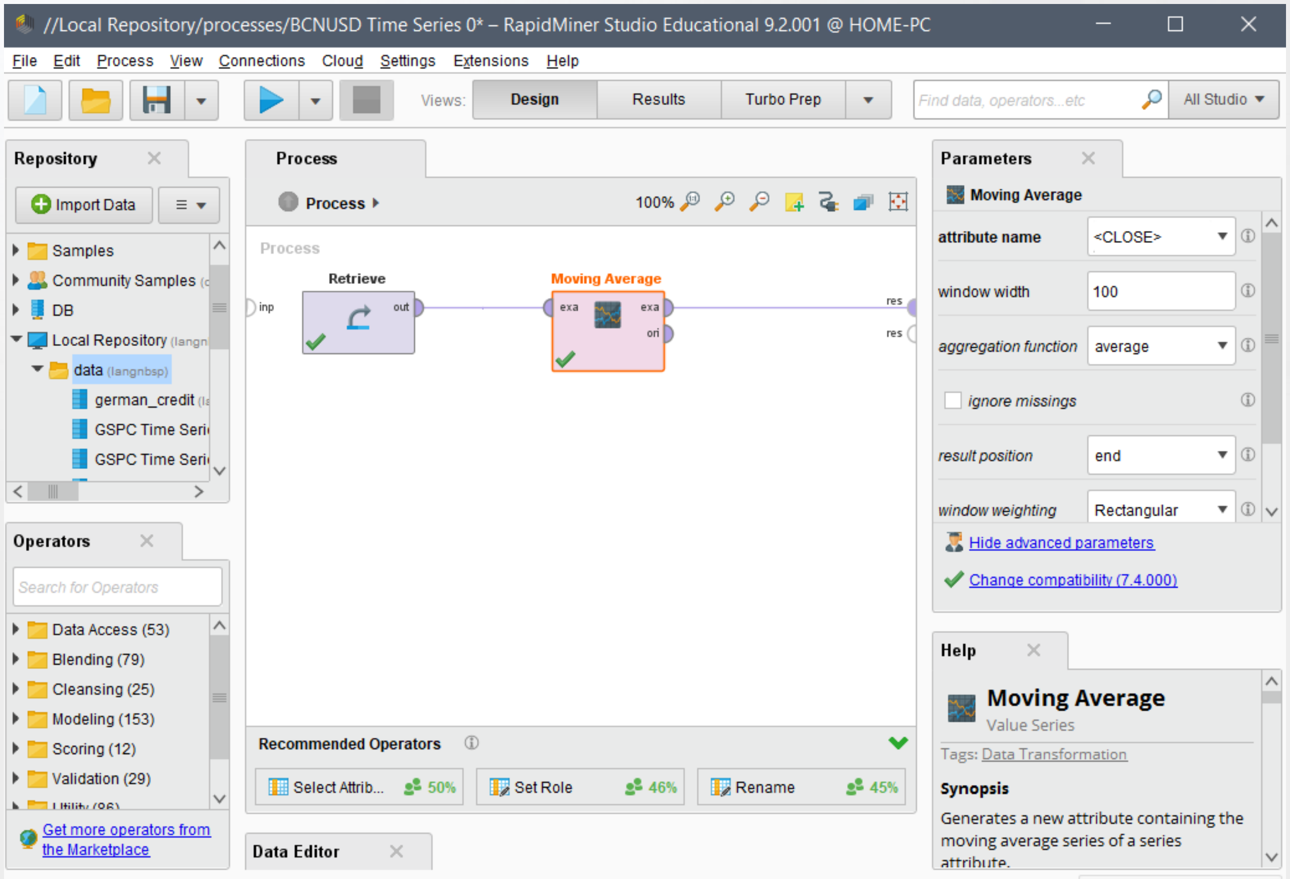
Research goals

- Analyze BTC/USD historical stock price data.
- Build NFF and LSTM neural network for price trend prediction.
- Transform initial data set to choose best data form for price trend prediction.
- Evaluate the applicability of the developed models.

Initial data set

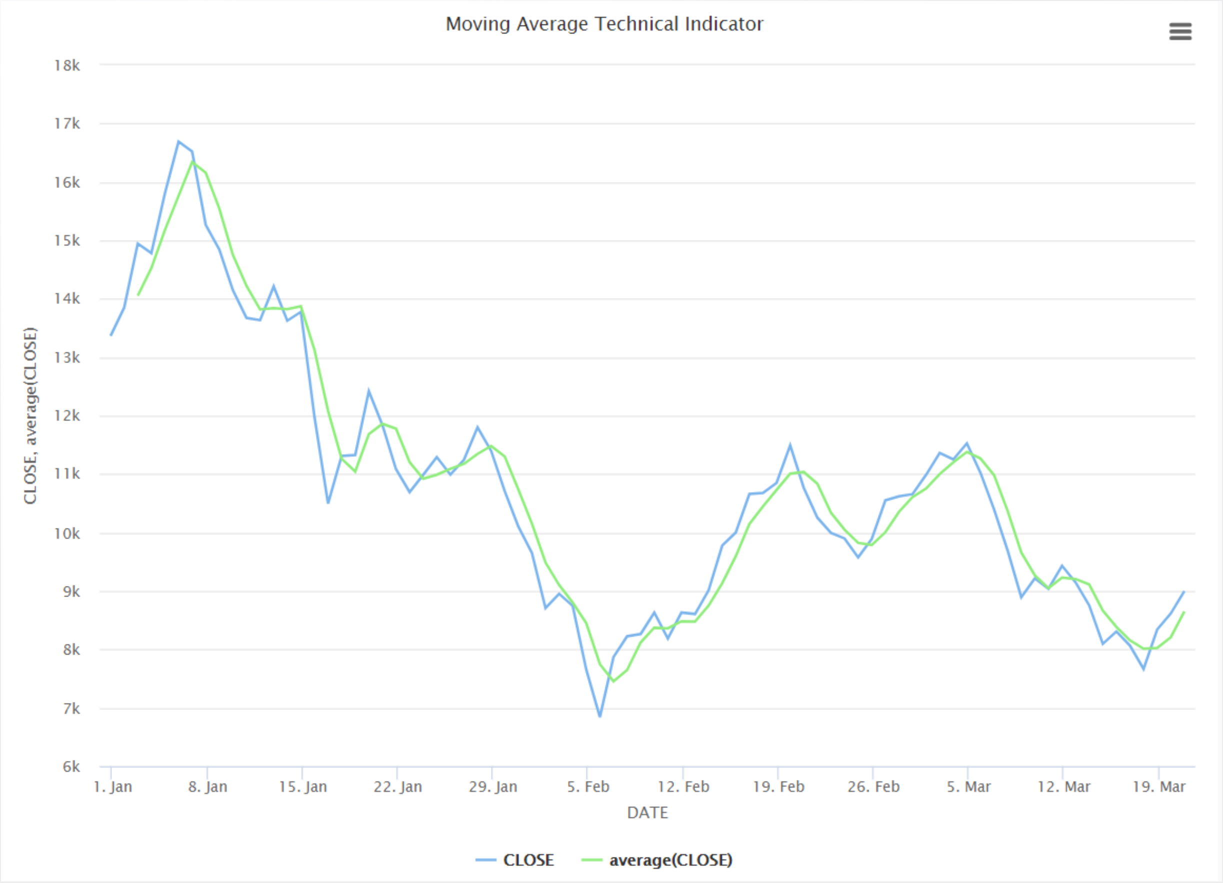


Data Discovery methods in RapidMiner: Moving Average



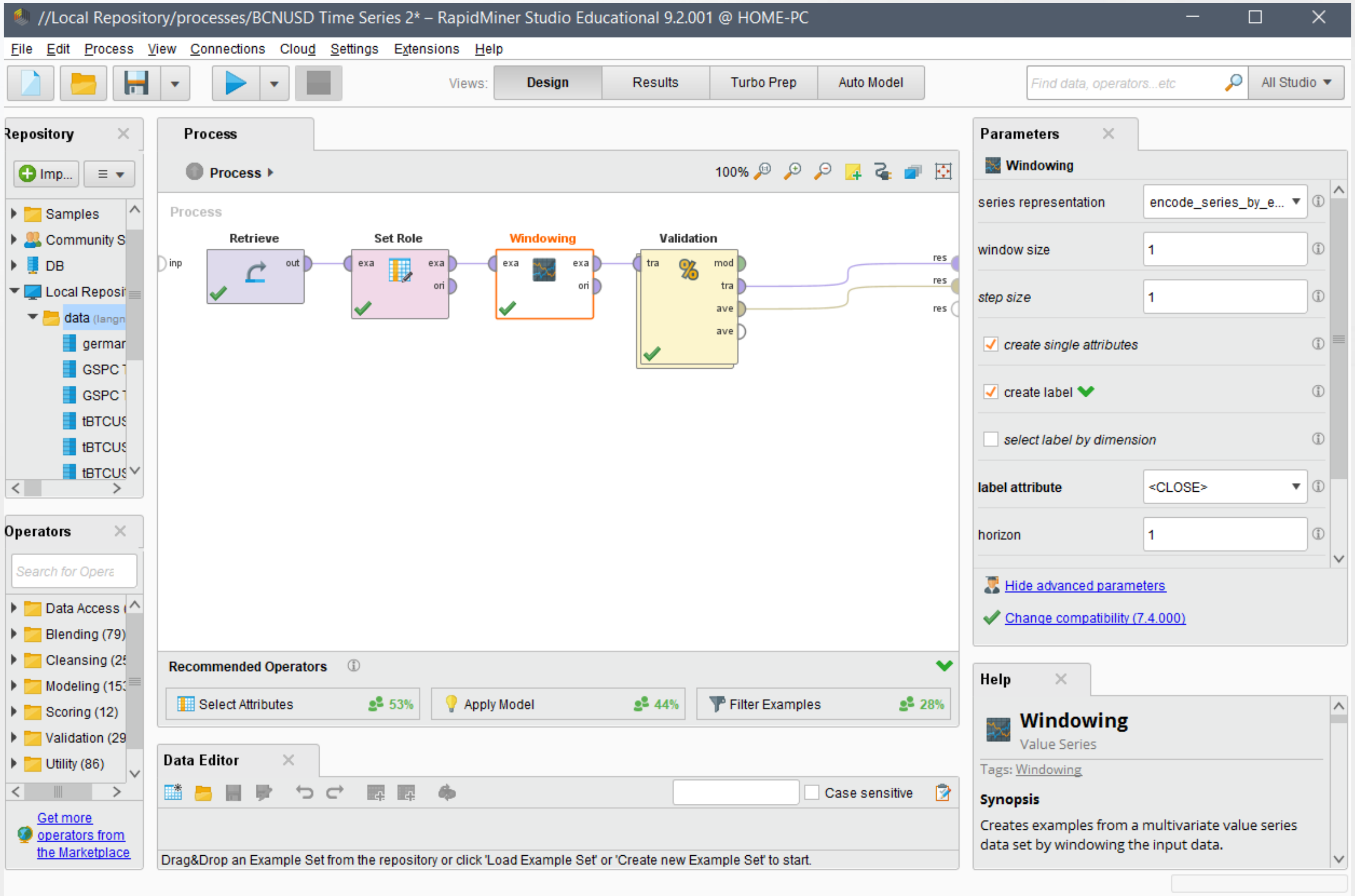
A moving average (MA) is a widely used indicator in technical analysis that helps smooth out price action by filtering out the "noise" from random short-term price fluctuations

Data Discovery methods in RapidMiner: Moving Average



When the price rises above MA - a buy signal occurs, when the price galls below MA - sell signal occurs

Financial time series modeling: feedforward neural network (FNN)



Windowing creates examples from a multivariate value series data set by windowing the input data.

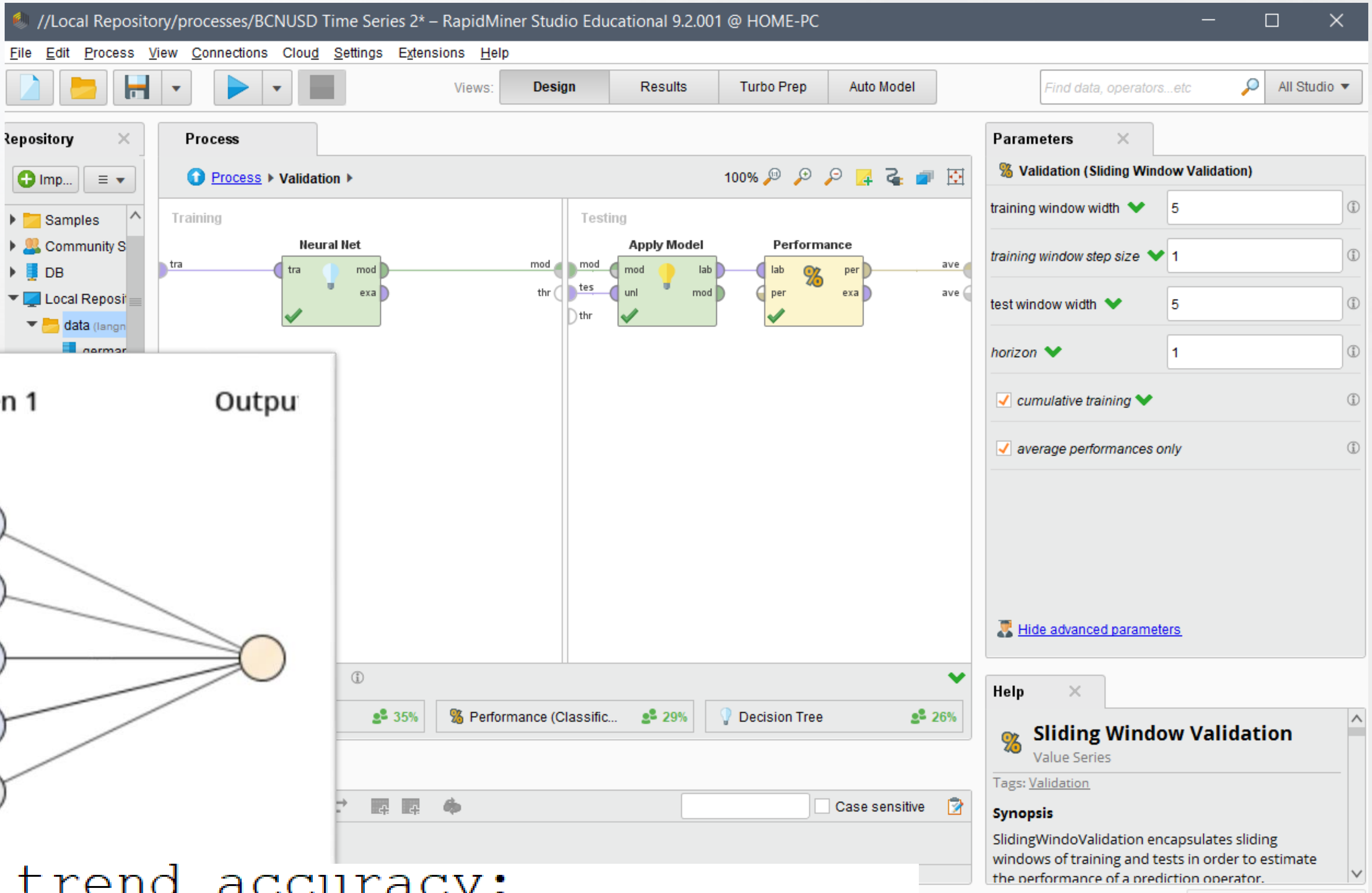
“By days” data input

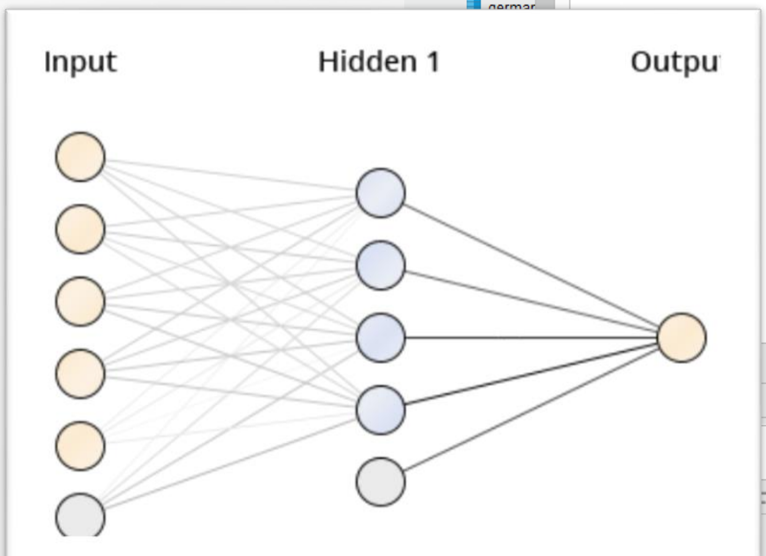
Row No.	DATE	label	CLOSE-0	OPEN-0	HIGH-0	LOW-0	VOLUME-0
1	Jan 1, 2018	13847.885	13364.234	13364.492	13378.431	13350.747	20
2	Jan 2, 2018	14944.656	13847.885	13847.038	13863.079	13832.069	38
3	Jan 3, 2018	14781.281	14944.656	14944.325	14958.738	14930.495	26
4	Jan 4, 2018	15808.432	14781.281	14781.252	14796.326	14766.950	31
5	Jan 5, 2018	16690.191	15808.432	15807.338	15822.295	15793.559	37
6	Jan 6, 2018	16521.085	16690.191	16689.948	16703.367	16677.782	20
7	Jan 7, 2018	15266.238	16521.085	16521.574	16534.645	16508.949	20
8	Jan 8, 2018	14842.637	15266.238	15267.078	15285.880	15247.283	44
9	Jan 9, 2018	14146.136	14842.637	14843.009	14857.991	14827.642	29
10	Jan 10, 2018	13672.237	14146.136	14145.615	14163.886	14128.943	39
11	Jan 11, 2018	13635.553	13672.237	13673.242	13694.620	13651.872	49
12	Jan 12, 2018	14211.244	13635.553	13635.253	13651.639	13620.113	25
13	Jan 13, 2018	13625.164	14211.244	14211.017	14223.393	14199.442	19
14	Jan 14, 2018	13776.550	13625.164	13625.611	13639.464	13611.815	24
15	Jan 15, 2018	11980.820	13776.550	13776.432	13789.084	13764.702	22
16	Jan 16, 2018	10494.780	11980.820	11982.304	12015.301	11949.857	91
17	Jan 17, 2018	11309.774	10494.780	10494.668	10530.102	10460.131	91
18	Jan 18, 2018	11326.263	11309.774	11309.694	11333.838	11286.630	58
19	Jan 19, 2018	12417.673	11326.263	11326.107	11344.212	11308.639	33
20	Jan 20, 2018	11837.638	12417.673	12416.778	12431.195	12402.576	31
21	Jan 21, 2018	11084.196	11837.638	11838.533	11853.596	11823.217	36
22	Jan 22, 2018	10692.331	11084.196	11084.851	11102.415	11066.876	45
23	Jan 23, 2018	10986.839	10692.331	10692.259	10710.087	10675.475	43

ExampleSet (67 examples, 2 special attributes, 5 regular attributes)

Label – training value of <CLOSE>, explained by other fields

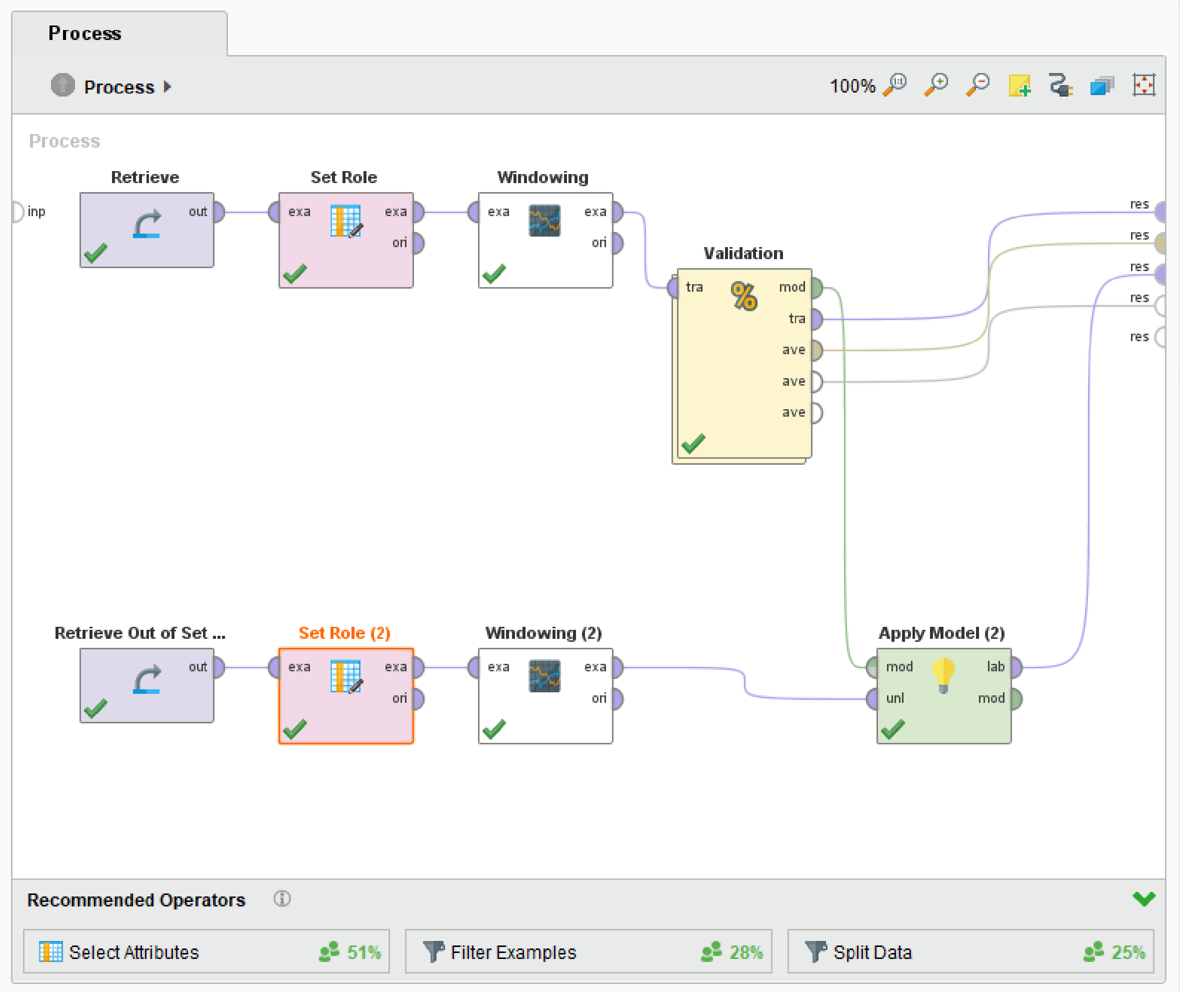
Build FNN model






```
prediction_trend_accuracy:
0.505 +/- 0.147 (micro average: 0.505)
```

Training and applying FNN model



The process describes FNN model and uses training / test data to form prediction

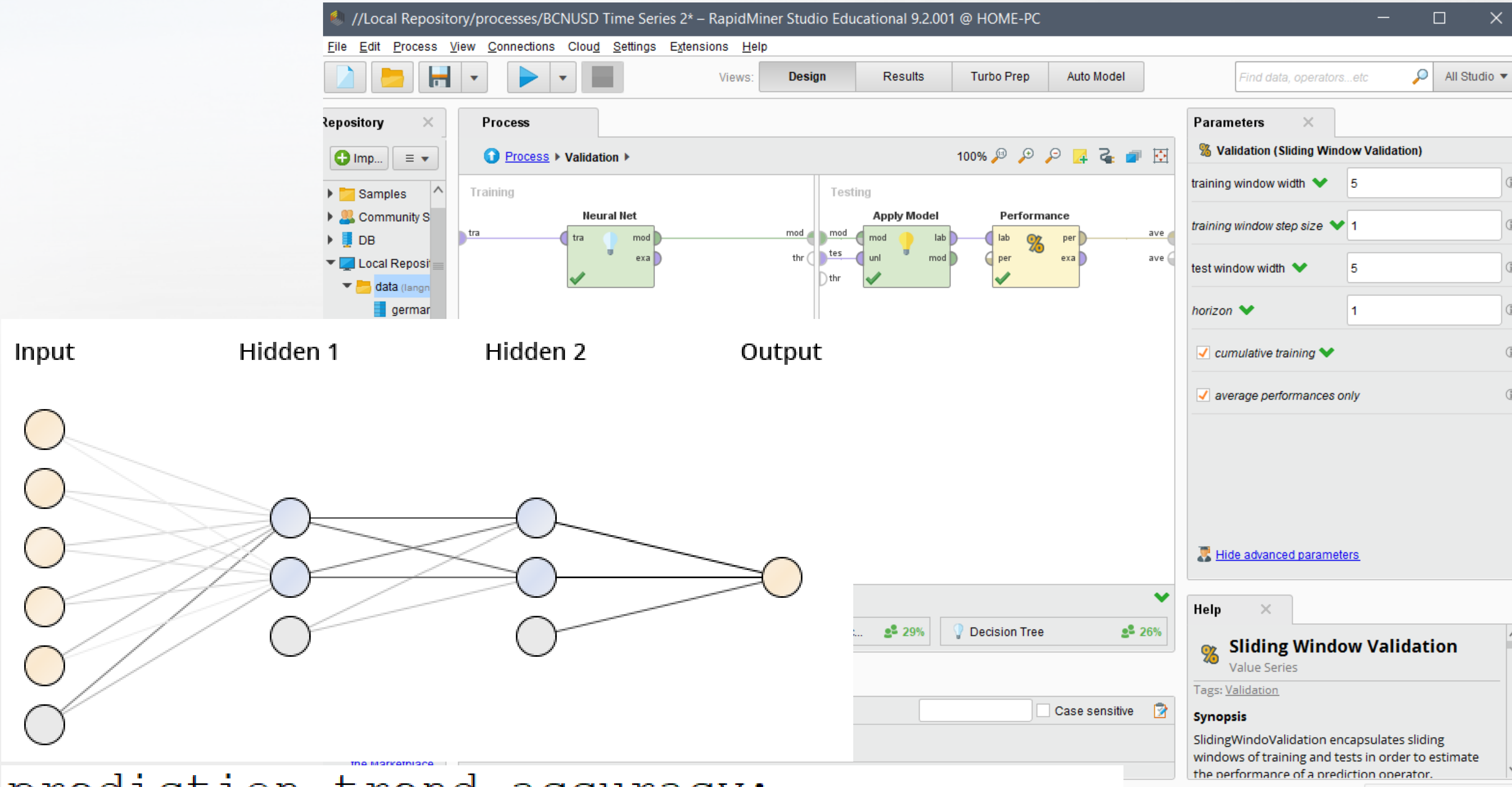
“By days Data” FFN applying results

Row No.	DATE	prediction(label)	CLOSE-0	OPEN-0	HIGH-0	LOW-0	VOLUME-0
1	Mar 10, 2018	9360.334	9217.467	9217.764	9226.324	9208.907	35
2	Mar 11, 2018	9149.105	9036.514	9035.996	9046.232	9026.055	48
3	Mar 12, 2018	9482.376	9431.860	9432.252	9440.830	9423.264	46
4	Mar 13, 2018	9267.550	9147.008	9147.055	9156.589	9137.400	42
5	Mar 14, 2018	8901.744	8753.489	8754.141	8762.067	8744.903	54
6	Mar 15, 2018	8425.080	8096.152	8096.068	8105.473	8087.087	56
7	Mar 16, 2018	8651.105	8306.558	8306.659	8314.106	8299.534	38
8	Mar 17, 2018	8507.647	8065.510	8065.815	8072.156	8059.483	33
9	Mar 18, 2018	8128.289	7666.010	7665.741	7675.343	7656.833	61
10	Mar 19, 2018	8619.864	8340.927	8340.596	8349.466	8332.866	50
11	Mar 20, 2018	8879.925	8616.058	8615.805	8622.753	8609.461	37
12	Mar 21, 2018	9228.180	8997.493	8997.453	9002.975	8992.341	26

Date	Direction	Correctness
10.03.18	-1	1
11.03.18	1	1
12.03.18	-1	1
13.03.18	-1	1
14.03.18	-1	1
15.03.18	-1	1
16.03.18	1	1
17.03.18	1	0
18.03.18	1	0

FNN model predicts the **price direction** on day close

FNN model applaying on full data




```
prediction_trend_accuracy:
0.642 +/- 0.136 (micro average: 0.642)
```

Full Data FFN applying results

Row No.	FPERIOD	prediction(la...	FCLOSE-0	FOPEN-0	FMAX-0	FLOW-0	FVOLUME-0
1	84958	10324.595	10307	10314	10316	10305	19
2	84959	10314.938	10296	10307	10307	10295	13
3	84960	10312.818	10300	10296	10306	10290	16
4	84961	10311.646	10298.494	10300	10300	10290	12
5	84962	10293.923	10266	10297	10297	10266	88
6	84963	10258.850	10226	10266	10266	10210	347
7	84964	10258.125	10257	10228	10258	10223	60
8	84965	10255.356	10228	10258	10261.918	10227	29
9	84966	10240.200	10225	10227	10228	10222	14
10	84967	10246.449	10235	10225	10240	10224.124	30
11	84968	10262.134	10257	10235	10258	10234	23
12	84969	10270.546	10256	10255	10260	10252	4
13	84970	10275.699	10265	10256	10265	10255	17
14	84971	10278.332	10265	10264	10265	10259.480	21
15	84972	10273.456	10255	10264	10265	10254	16
16	84973	10272.571	10261	10255	10264	10250	13
17	84974	10283.792	10272	10265	10272	10265	6
18	84975	10299.373	10295	10272	10295	10271	14
19	84976	10309.534	10294	10295	10296	10294	12
20	84977	10298.781	10280	10294	10294	10274	17
21	84978	10303.642	10295	10281	10295	10281	10
22	84979	10312.442	10300	10295	10300	10294	16
23	84980	10323.636	10315	10301	10315	10300	35
24	84981	10342.158	10335	10315	10339	10314	54
25	84982	10354.394	10343	10336	10343	10335	10

ExampleSet (29,485 examples, 2 special attributes, 5 regular attributes)

Date	Direction	Correctness
84958	1	0
84959	1	1
84960	1	0
84961	-1	1
84962	-1	1
84963	1	1
84964	-1	1
84965	1	0
84966	1	1

Logarithmic return calculation

Logarithmic return formula:

$$R = \ln\left(\frac{V_f}{V_i}\right)$$

Logarithmic rate of return:

$$r_{\log} = \frac{\ln\left(\frac{V_f}{V_i}\right)}{t}$$

$$V_f = V_i e^{rt}$$

Extended input data set

Let FIELDS = [OPEN, HIGH, LOW , CLOSE, VOL]
Then:

$$FIELDS[j]_2 = LN\left(\frac{FIELDS[j]_{i_2}}{FIELDS[j]_{i_1}}\right)$$

$$FIELDS[j]_1 = LN\left(\frac{FIELDS[j]_{i_3}}{FIELDS[j]_{i_2}}\right)$$

$$FIELDS[j] = LN\left(\frac{FIELDS[j]_{i_4}}{FIELDS[j]_{i_3}}\right)$$

Where:
i – time periods,
j – fields index.

Extended input data set

DATE	TIME	OPEN-2	HIGH-2	LOW-2	CLOSE-2	VOL-2	OPEN-1	HIGH-1	LOW-1	CLOSE-1	VOL-1	OPEN	HIGH	LOW	CLOSE	VOL	TARGET	indicator
20180101	400	-0.00016	-0.00016	-0.00092	-0.00092	22	-0.00087	-0.00087	-0.00108	-0.00077	20	-0.00077	-6.45E-06	0.000308	0	12	-0.00252	0
20180101	500	-0.00087	-0.00087	-0.00108	-0.00077	20	-0.00077	-6.45E-06	0.000308	0	12	-3.08E-07	-0.00077	-0.00105	-0.00072	9	-0.00294	0
20180101	600	-0.00077	-6.45E-06	0.000308	0	12	-3.08E-07	-0.00077	-0.00105	-0.00072	9	-0.00072	-0.00072	-0.00025	-0.00057	3	-0.00216	0
20180101	700	-3.08E-07	-0.00077	-0.00105	-0.00072	9	-0.00072	-0.00072	-0.00025	-0.00057	3	-0.00057	0.000108	-9.26E-05	-9.20E-05	6	-0.00151	0
20180101	800	-0.00072	-0.00072	-0.00025	-0.00057	3	-0.00057	0.000108	-9.26E-05	-9.20E-05	6	-9.23E-05	-0.00077	-0.00015	-0.00013	17	-0.00174	0
20180101	900	-0.00057	0.000108	-9.26E-05	-9.20E-05	6	-9.23E-05	-0.00077	-0.00015	-0.00013	17	-0.00012	-0.00012	-0.0002	-0.00023	6	-0.00102	0
20180101	1000	-9.23E-05	-0.00077	-0.00015	-0.00013	17	-0.00012	-0.00012	-0.0002	-0.00023	6	-0.00023	-0.00023	0	2.16E-06	8	0.001664	1
20180101	1100	-0.00012	-0.00012	-0.0002	-0.00023	6	-0.00023	-0.00023	0	2.16E-06	8	2.47E-06	0.002111	2.47E-06	0.002111	18	0.001744	1
20180101	1200	-0.00023	-0.00023	0	2.16E-06	8	2.47E-06	0.002111	2.47E-06	0.002111	18	0.002111	0	0.00208	-1.17E-05	7	0.000988	1
20180101	1300	2.47E-06	0.002111	2.47E-06	0.002111	18	0.002111	0	0.00208	-1.17E-05	7	-3.08E-05	-3.08E-05	-0.00086	-0.00088	4	0.000578	1
20180101	1400	0.002111	0	0.00208	-1.17E-05	7	-3.08E-05	-3.08E-05	-0.00086	-0.00088	4	-0.00086	-0.00086	-0.00085	-0.00064	8	0.000853	1
20180101	1500	-3.08E-05	-3.08E-05	-0.00086	-0.00088	4	-0.00086	-0.00086	-0.00085	-0.00064	8	-0.0006	0.000215	0.000254	0.000278	9	-0.00126	0
20180101	1600	-0.00086	-0.00086	-0.00085	-0.00064	8	-0.0006	0.000215	0.000254	0.000278	9	0.000232	-0.00058	0.000231	-1.54E-06	5	-0.00152	0
20180101	1700	-0.0006	0.000215	0.000254	0.000278	9	0.000232	-0.00058	0.000231	-1.54E-06	5	-1.23E-06	-1.23E-06	-0.00029	-0.00029	2	-0.00052	0
20180101	1800	0.000232	-0.00058	0.000231	-1.54E-06	5	-1.23E-06	-1.23E-06	-0.00029	-0.00029	2	-0.00029	-6.54E-05	6.17E-07	0.000141	7	0.000135	1
20180101	1900	-1.23E-06	-1.23E-06	-0.00029	-0.00029	2	-0.00029	-6.54E-05	6.17E-07	0.000141	7	0.000141	-3.08E-05	0.000141	5.52E-05	8	2.56E-05	1
20180101	2000	-0.00029	-6.54E-05	6.17E-07	0.000141	7	0.000141	-3.08E-05	0.000141	5.52E-05	8	5.52E-05	0.000123	5.52E-05	0.000123	5	0.000582	1
20180101	2100	0.000141	-3.08E-05	0.000141	5.52E-05	8	5.52E-05	0.000123	5.52E-05	0.000123	5	0.000123	0.000555	0.000123	0.000555	1	0.000936	1
20180101	2200	5.52E-05	0.000123	5.52E-05	0.000123	5	0.000123	0.000555	0.000123	0.000555	1	0.000555	9.24E-05	0.000555	6.13E-05	4	0.000794	1
20180101	2300	0.000123	0.000555	0.000123	0.000555	1	0.000555	9.24E-05	0.000555	6.13E-05	4	6.13E-05	-3.11E-05	6.01E-05	-9.24E-07	4	0.000711	1
20180101	2400	0.000555	9.24E-05	0.000555	6.13E-05	4	6.13E-05	-3.11E-05	6.01E-05	-9.24E-07	4	-1.23E-06	-9.24E-07	-2.71E-05	-2.71E-05	7	0.000278	1
20180101	2500	6.13E-05	-3.11E-05	6.01E-05	-9.24E-07	4	-1.23E-06	-9.24E-07	-2.71E-05	-2.71E-05	7	-2.68E-05	-2.71E-05	-0.00031	-0.00031	3	-0.00077	0
20180101	2600	-1.23E-06	-9.24E-07	-2.71E-05	-2.71E-05	7	-2.68E-05	-2.71E-05	-0.00031	-0.00031	3	-0.00031	-0.00031	-0.00049	-0.00049	2	-0.00126	0

Total of records: 114442

Extended input data set

Target fields:

$$TARGET_i = LN\left(\frac{CLOSE_{i_5}}{OPEN_{i_1}}\right)$$

$$INDICATOR_i = \begin{cases} 1, TARGET_i > 0 \\ 0, TARGET_i \leq 0 \end{cases}$$

Where:

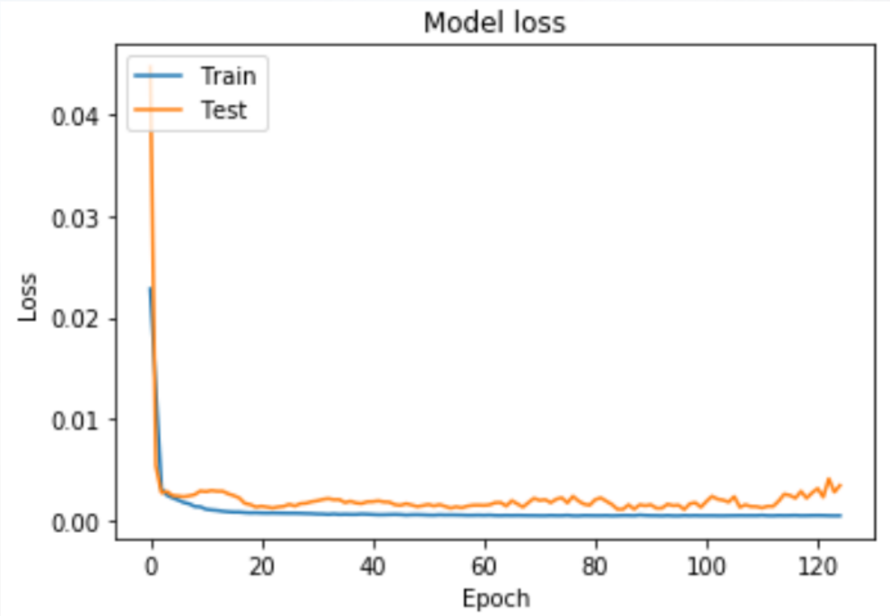
i – time periods.

Log Return modeling: Long short-term memory (LSTM)

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(20, 60, 100)	42400
dropout_1 (Dropout)	(20, 60, 100)	0
lstm_2 (LSTM)	(20, 60)	38640
dropout_2 (Dropout)	(20, 60)	0
dense_1 (Dense)	(20, 20)	1220
dense_2 (Dense)	(20, 1)	21

Layer Dense_1 activation function – ReLU (Rectified Linear Unit)
Layer Dense_2 activation function – Sigmoid

Log Return modeling: Long short-term memory (LSTM)




```
prediction_trend_accuracy:
0.817 +/- 0.128 (micro average: 0.817)
```

```
Building model...
Train on 91480 samples, validate on 11400 samples
Epoch 1/300
 - 970s - loss: 0.0228 - val_loss: 0.0448
Epoch 2/300
 - 969s - loss: 0.0130 - val_loss: 0.0054
Epoch 3/300
 - 944s - loss: 0.0031 - val_loss: 0.0027
Epoch 4/300
 - 848s - loss: 0.0024 - val_loss: 0.0028
Epoch 5/300
 - 799s - loss: 0.0022 - val_loss: 0.0024
Epoch 6/300
 - 771s - loss: 0.0020 - val_loss: 0.0024
Epoch 7/300
 - 808s - loss: 0.0017 - val_loss: 0.0023
Epoch 8/300
 - 759s - loss: 0.0016 - val_loss: 0.0024
Epoch 9/300
 - 764s - loss: 0.0018 - val_loss: 0.0025
Epoch 10/300
 - 798s - loss: 0.0023 - val_loss: 0.0029
Epoch 11/300
 - 804s - loss: 0.0016 - val_loss: 0.0028
Epoch 12/300
 - 912s - loss: 0.0019 - val_loss: 0.0029
 ...
Epoch 125/300
 - 975s - loss: 4.2855e-04 - val_loss: 0.0034
Epoch 00125: early stopping
saving model...
```

Results comparison

Model type	Data size	Prediction accuracy	Training time*
FNN (days as time period)	80	50%	24 min
FNN (minutes as time period)	114442	64%	4h 12 min
LSTM (Log return indicator)	114442	82%	23h 18 min

* Hardware configuration:
Intel® Core™ i7-3630QM CPU @ 2.49Ghz
RAM 16.0 Gb

Acknowledgements

This work was supported by the Russian Foundation for Basic Research under Grant No 18-01-00910