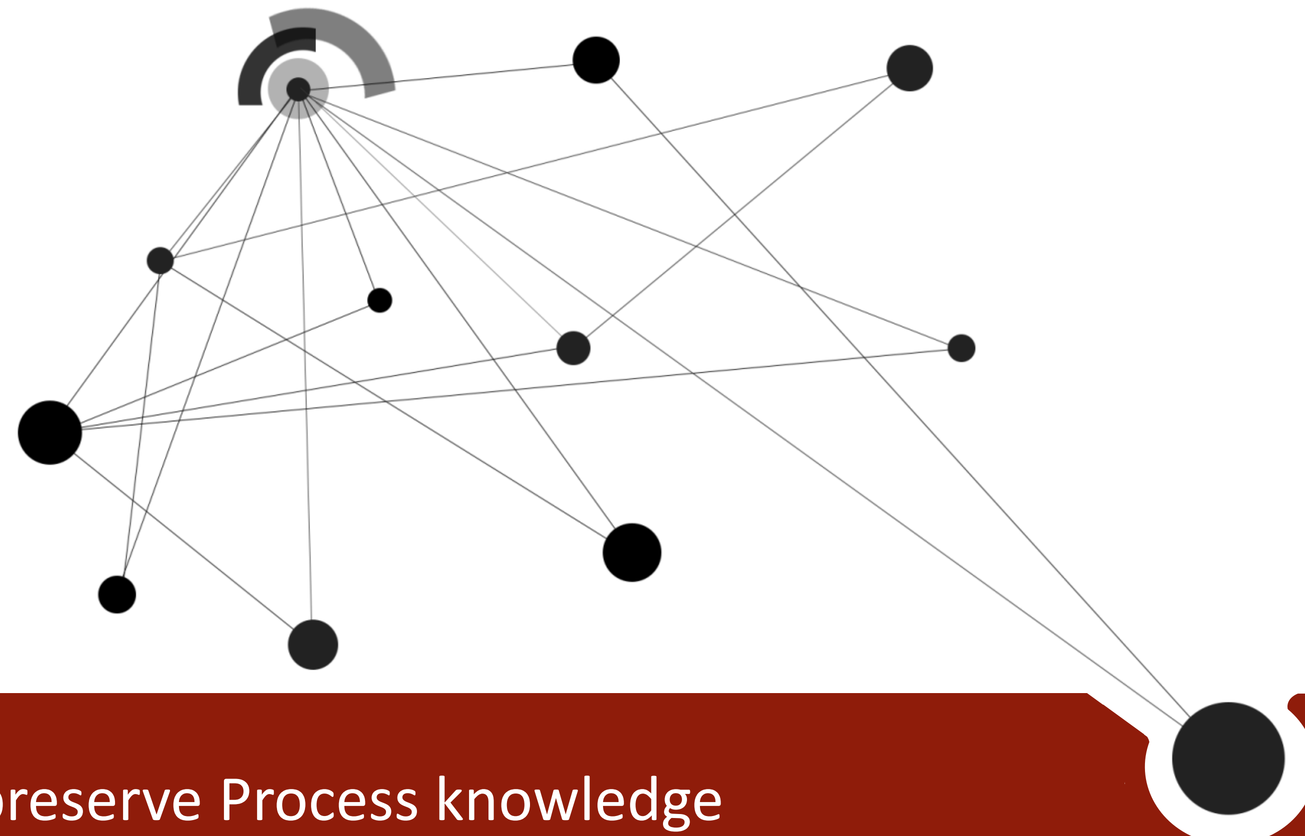


Generate and preserve Process knowledge

Control and optimize with Predictive Analytics
and knowledge-based Systems

J. Gottschling, Universität Duisburg Essen
D. Hartmann, Hochschule Kempten

Predictive Analytics – What is it?

Predictive Analytics uses methods from the fields of *Statistics*, *Machine Learning*, and *Data Mining* to give predictions based on current and historical data.

The above mentioned fields are not well seperated. Data mining methods covers e.g. Classification (Clustering), Decision trees and Association Analyzes.

Link: http://www.computerwoche.de/a/was-ist-was-bei-predictive-analytics,3098583

Predictive policing is about the prediction of crimes. The software gives predictions based on past years data (e.g., location, time, Prosecutions) on criminal offenses in a certain region. The police patrol cars can then be concentrated in this vulnerable region (See http://www.computerwoche.de/a/was-ist-was-bei-predictive-analytics,3098583).

Google Maps – An example to get started

Static Route planners calculate the arrival time with an average Speed and the Speed profile of the route.

But: The arrival time depends significantly on the day of the week and the time.

Whether the trip from Duisburg to Dortmund on the A40 at 7:00 am begins on Wednesday or on Sunday, has a significant impact on the Arrival time.

Dynamic route planners consider - in addition to the average Speed and the Speed profile, ***historical and current data*** (existing congestion and construction sites) in the calculation of the route and the arrival time.

Such data is obtained e.g. by Floating Car Data / Floating Phone Data.

Google Maps Product manager Dave Barth:
"When we summarize your speed with the speed of other smartphones on the road, from thousands of phones, moving around the streets of any city at any time, we get a good Picture of current Traffic conditions."

The basis of supervised Machine Learning is a set of similar datasets. A dataset is a group of related data fields, e.g. *T, C, Si, Mg* and *Strain*. If the datasets are stored in a table, a dataset is one row of this table.

T [°C]	C [%]	Si [%]	Mg [%]	Dehnung [%]
1.381,23	3,59	2,71	0,052	20,0
1.385,33	3,56	2,72	0,041	20,3
1.381,40	3,61	2,63	0,042	7,6
1.385,40	3,62	2,68	0,049	19,3
1.379,00	3,63	2,66	0,054	20,0
1.380,00	3,62	2,63	0,053	19,3
1.379,50	3,72	2,78	0,048	20,0
1.379,20	3,67	2,72	0,055	15,9
1.380,68	3,58	2,62	0,047	17,6

The fields (in the following also called variables) of such a Dataset are distinguished into functionally *independent* and functionally *dependent* fields. In the above example, *T, C, Si,* and *Mg* are the functionally independent variables, and the *Strain* is a functionally dependent variable. Several functionally dependent variables are also possible.

The aim of Machine Learning:

To predict the functionally *dependent* variables "sufficiently good" from the functionally *independent* variables on the basis of training data.

The variables of measured Datasets are *Random variables* in the stochastic sense since they are subject to random fluctuations, and for these random variables, the terms *stochastic dependence* or *independence* are well defined.

Therefore, the terms *functional dependence* and *independence* are used for the measured values.

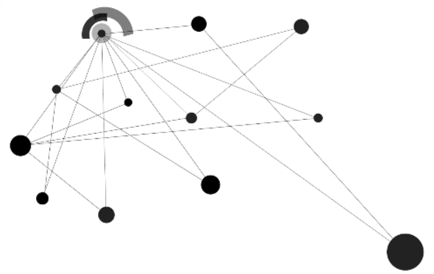
Functionally dependent Random variables ($A_1, \dots , A_k$) with a Prediction function F are predicted from the functionally independent random variables ($U_1, \dots, U_n$) .
Thus: $(A_1, \dots , A_k) = F(U_1, \dots, U_n)$.

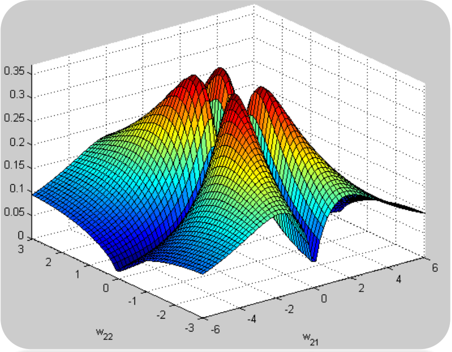
These Prediction functions are trained and validated on the basis of Measured data.

There are many methods of Machine Learning that can generate such Prediction functions:

- Neural Networks (NN),
- Bayesian Networks (BN),
- K Nearest Neighbors (KNN)
- Decision Trees (DT),
- Multiple Regression (MR),
- Support Vector Machines (SVM),
- Logistic Regression (LOR)

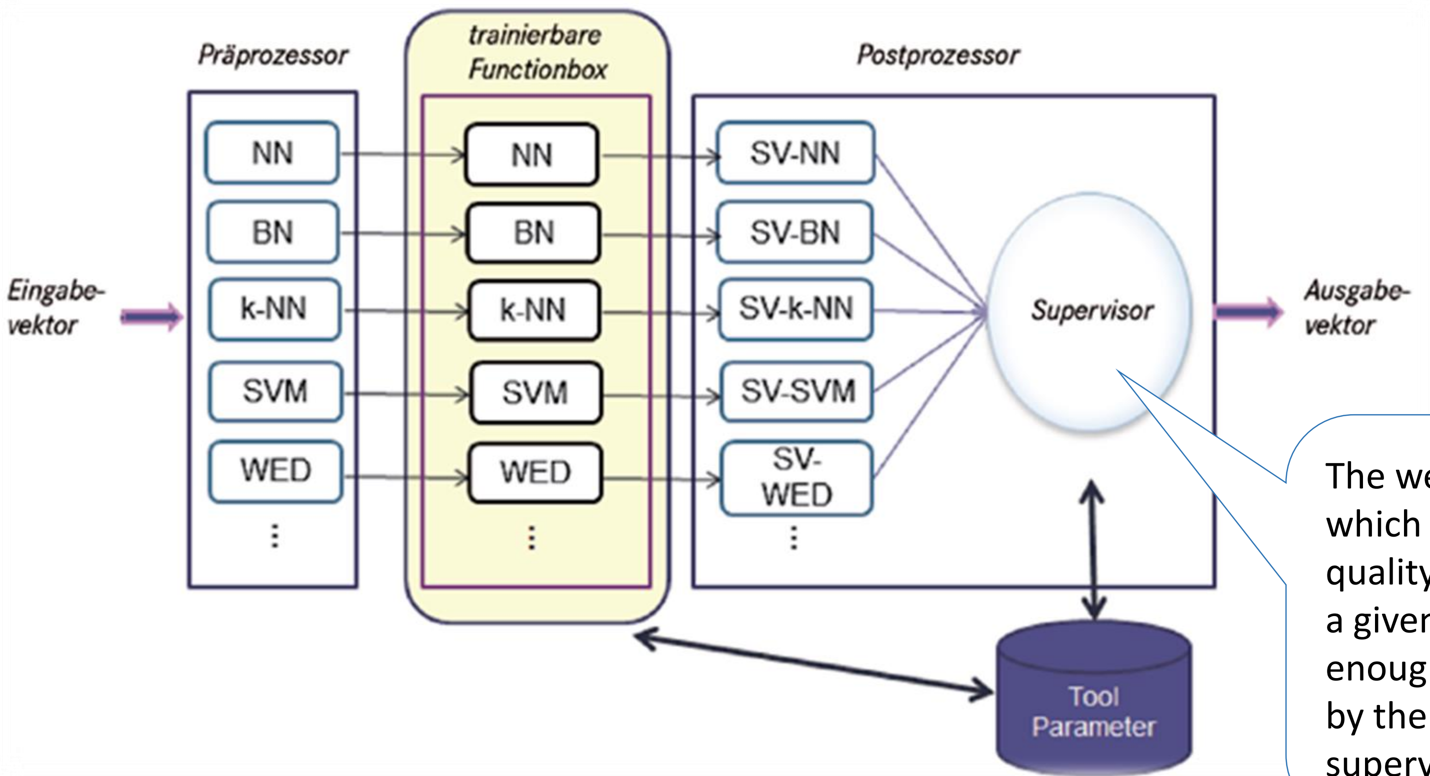
...





Predictive Analytics with EIDOminer

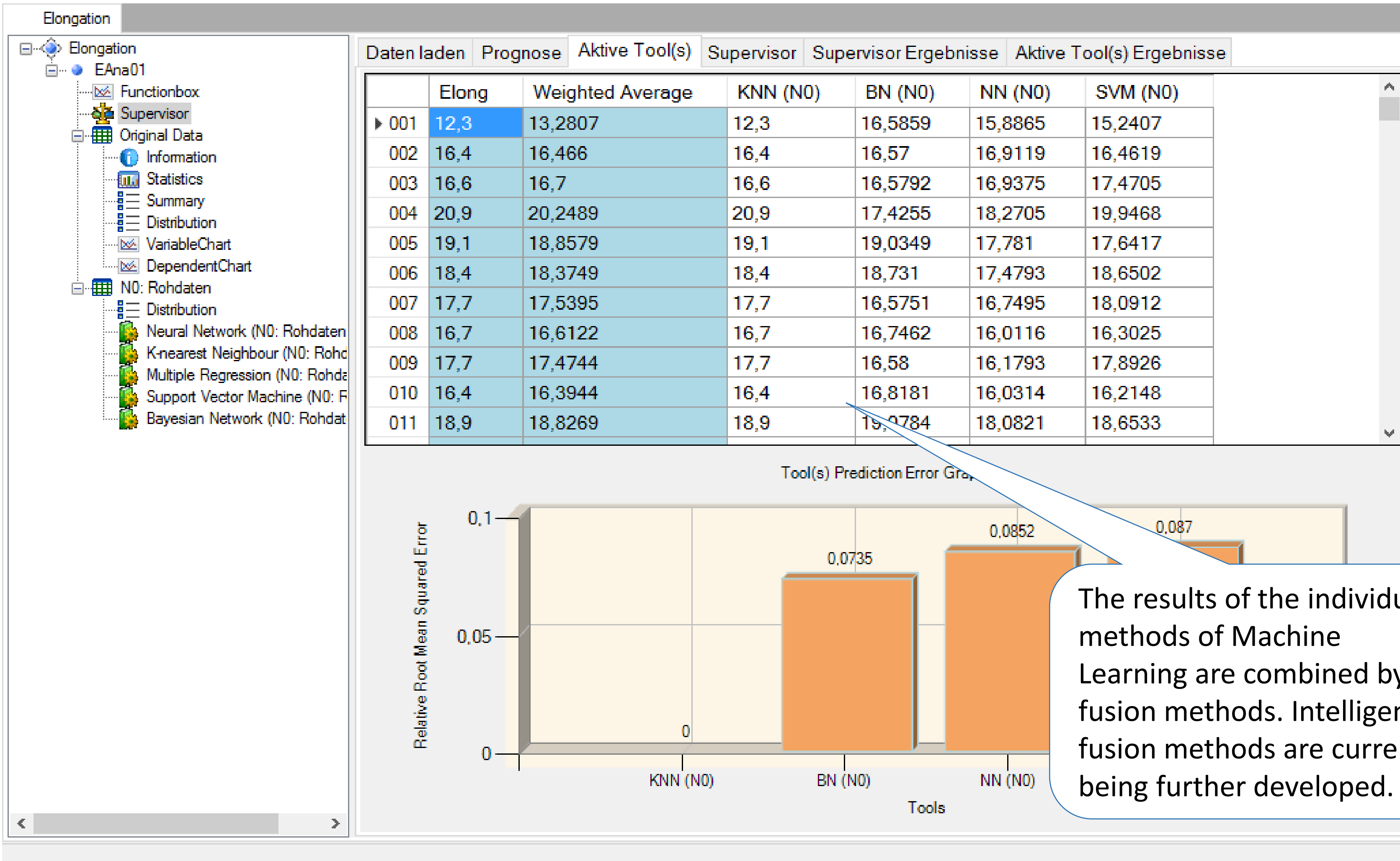
The Intelligent Analysis Manager is the core module of the *EIDOminer* and summarizes several methods of Machine Learning in a *Functionbox*.



The weaker methods, which can not predict the quality characteristics of a given process well enough, are switched off by the learning supervisor.

Predictive Analytics with EIDOminer

Hochschule Kempten
UNIVERSITÄT DUISBURG ESSEN *Offen im Denken*



	Elong	Weighted Average	KNN (N0)	BN (N0)	NN (N0)	SVM (N0)
001	12,3	13,2807	12,3	16,5859	15,8865	15,2407
002	16,4	16,466	16,4	16,57	16,9119	16,4619
003	16,6	16,7	16,6	16,5792	16,9375	17,4705
004	20,9	20,2489	20,9	17,4255	18,2705	19,9468
005	19,1	18,8579	19,1	19,0349	17,781	17,6417
006	18,4	18,3749	18,4	18,731	17,4793	18,6502
007	17,7	17,5395	17,7	16,5751	16,7495	18,0912
008	16,7	16,6122	16,7	16,7462	16,0116	16,3025
009	17,7	17,4744	17,7	16,58	16,1793	17,8926
010	16,4	16,3944	16,4	16,8181	16,0314	16,2148
011	18,9	18,8269	18,9	19,0784	18,0821	18,6533

The results of the individual methods of Machine Learning are combined by fusion methods. Intelligent fusion methods are currently being further developed.

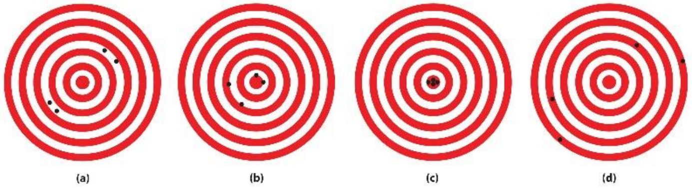
The predictions are only as good as the data on which they are based!

The software *EIDOminer* therefore examines in the *Preprocessor* whether the *functionally independent* variables from which the *functionally dependent* variables are to be predicted have statistical dependencies and whether the number of these *functionally independent* variables can be reduced (Dimension reduction).

If the functionally independent variables are correlated (not only linear), the measurement of their isolated influence on the functional dependent variable becomes more difficult. A variable, which is correlated with other variables, measures in part the influence of the other variables.

This so-called *Multicollinearity* can be eliminated or reduced by different methods.

Through Dimensional reduction such problems are corrected and the model becomes "slimmer", the model interpretation more precise.



If the functional dependent random variable is discrete, one speaks of *Classification* and of *Regression,* if it is continuous.

So:
Classification is the prediction of a class (e.g., OK, Critical, NONE), and *Regression* is the prediction of a real number.

But not only the Dependencies of the functionally independent variables are important.

Important is also the Question:

What is the effect of the functionally independent variables on the functionally dependent variables?

The Preprocessor, however, can not determine whether the set of functionally independent variables is complete. Complete in the sense that the independent variables can predict the dependent variable (s) well.

By increasing additional independent variables, the Prediction quality can often be significantly improved.

The Preprocessor includes the following pedictability function:

It will search for all Datasets whose functional independent variables are "close" to each other, while the functionally dependent variables are "wide" apart.

Such datasets can not be processed by the Machine Learning algorithms.

A new added variable often provides a solution.

Hochschule Kempten
University of Applied Sciences

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

If the data is excellent, the Predictions are also excellent.

As an example, consider the dataset

TestData01 (100 lines):
6 functionally independent variables **I1 - I6** and a functionally dependent variable **D**

I1 - randomly generated with a Normal distribution
I2 - 3*I1 + Noise
I3 - 5*I1 + Noise
I4 - randomly generated with a Normal distribution
I5 - I1 + I4 + Noise
I6 - π*I4 + Noise
D - 2*I1^0,5 + 3 * I4^1.2

J2

	A	B	C	D	E	F	G
1	**I1**	**I2**	**I3**	**I4**	**I5**	**I6**	**D**
2	22,63	67,88	113,17	30,51	53,15	95,86	190,837
3	18,13	54,4	90,65	28,05	46,17	88,12	172,439
4	17,66	52,98	88,31	27,71	45,38	87,04	169,947
5	16,83	50,48	84,11	31,1	47,93	97,7	193,743
6	21,24	63,76	106,22	33,13	54,38	104,09	209,382
7	26,97	80,92	134,87	26,61	53,58	83,63	164,264
8	21,82	65,46	109,1	29,1	50,92	91,42	180,656
9	18,07	54,19	90,32	29,83	47,91	93,72	184,986
10	22,09	66,26	110,45	26,02	48,08	81,73	159,192
11	17,18	51,52	85,9	31,68	48,85	99,55	197,988
12	16,44	49,31	82,19	29,17	45,61	91,64	179,918
13	21,79	65,37	108,97	33,82	55,61	106,26	214,513
14	18,06	54,17	90,27	29,22	47,27	91,81	180,661
15	20,79	62,37	103,93	30,82	51,59	96,83	192,655
16	23,52	70,57	117,59	27,72	51,22	87,08	171,311
17	21	63,02	105,03	28,75	49,75	90,31	178,01
18	22,32	66,97	111,64	27,66	49,98	86,91	170,641
19	15,99	47,95	79,91	30,79	46,76	96,72	191,319

Predictive Analytics with EIDOminer

I1	I2	I3	I4	I5	I6
22,63	67,88	113,17	30,51	53,15	95,86
18,13	54,4	90,65	28,05	46,17	88,12
17,66	52,98	88,31	27,71	45,38	87,04
16,83	50,48	84,11	31,1	47,93	97,7
21,24	63,76	106,22	33,13	54,38	104,09
26,97	80,92	134,87	26,61	53,58	83,63
21,82	65,46	109,1	29,1	50,92	91,42
18,07	54,19	90,32	29,83	47,91	93,72

Korrelationsmatrix

	I1	I2	I3	I4	I5	I6
I1	1,000000	0,999998	0,999999	-0,100961	0,765300	-0,100512
I2	0,999998	1,000000	0,999999	-0,100874	0,765358	-0,100425
I3	0,999999	0,999999	1,000000	-0,100679	0,765483	-0,100231
I4	-0,100961	-0,100874	-0,100679	1,000000	0,563109	0,999998
I5	0,765300	0,765358	0,765483	0,563109	1,000000	0,563481
I6	-0,100512	-0,100425	-0,100231	0,999998	0,563481	1,000000

fx =KORREL(A$3:A$102;$B$3:$B$102)

According to the above linear correlation matrix, only the columns I1, I4 and I5 should be used to train the tools of the Machine Learning.

The *EIDOminer* software detects the correlated variables when loading the dataset and suggests removing these variables. I5 = I1 + I4 + Noise is not recognized because the correlation coefficient Is set to 90% by default.

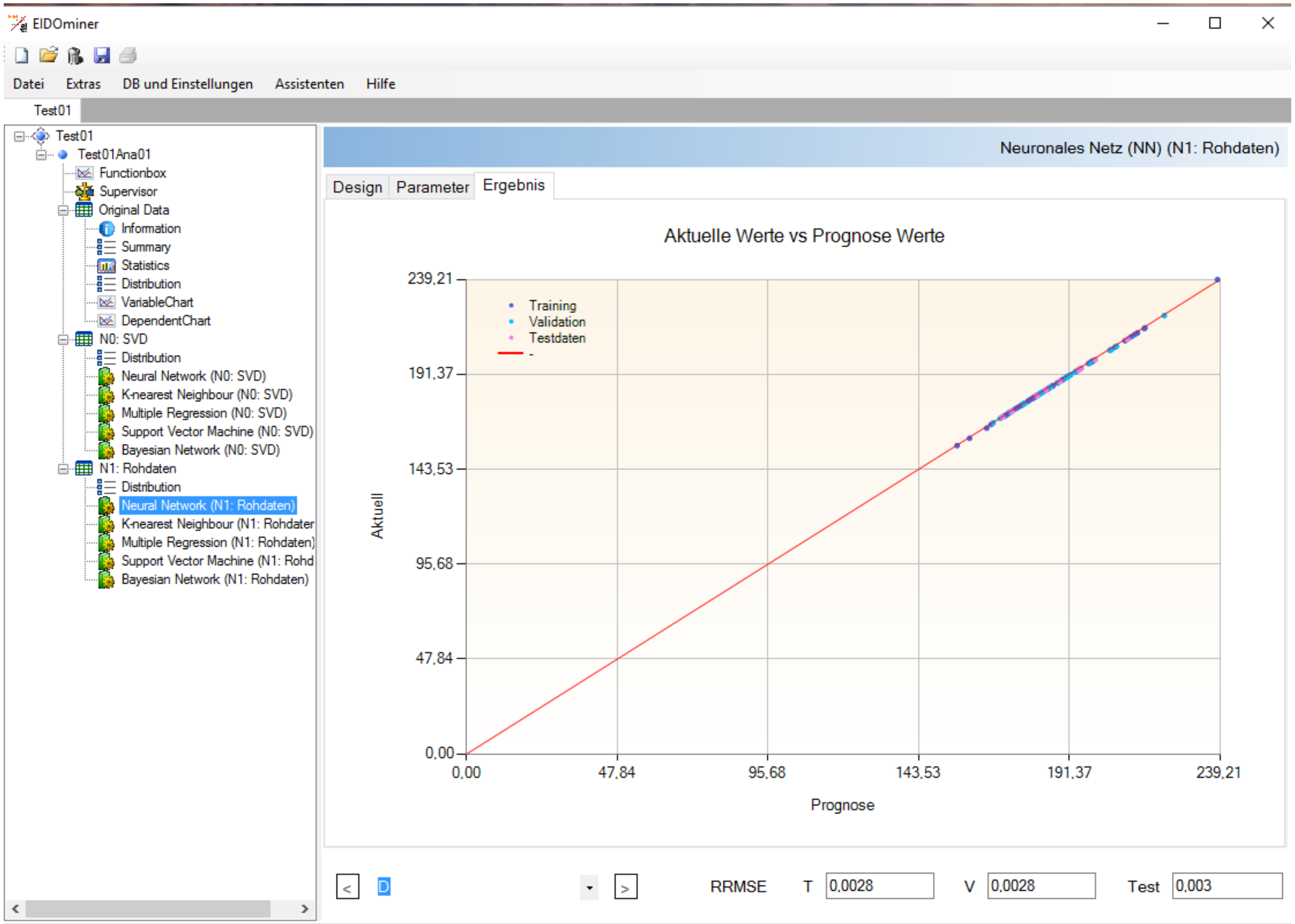
Korrelierende Variablen

Entfernen	Variable	Korrelation	PP-Tool-Name
☑	I2	I1, I2=99,9998%	N0: Rohdaten
☑	I3	I1, I3=99,9999%	N0: Rohdaten
☑	I3	I2, I3=99,9999%	N0: Rohdaten
☑	I6	I4, I6=99,9998%	N0: Rohdaten

Entfernen | Weiter ohne entfernen

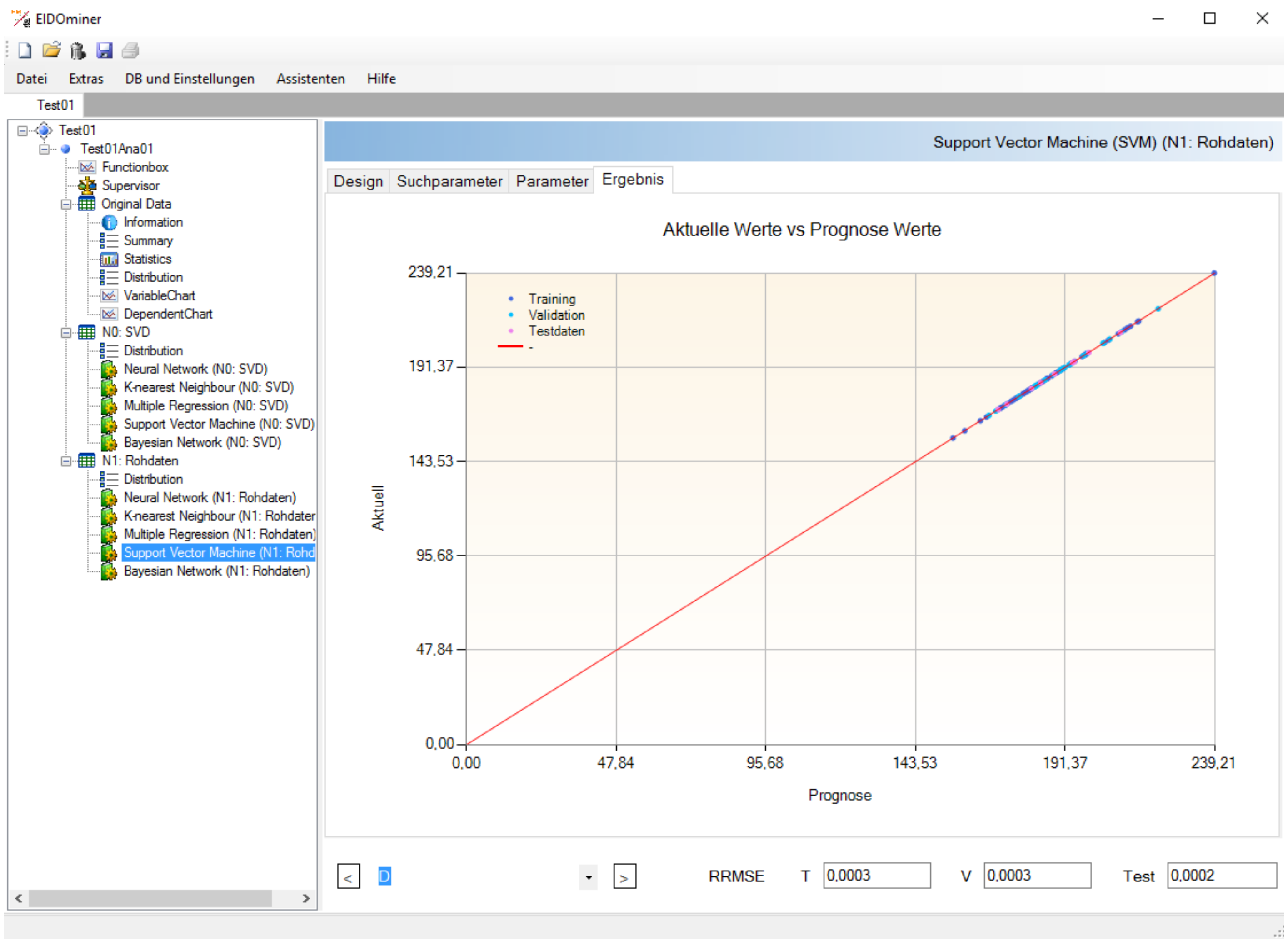
Predictive Analytics with EIDOminer

Result Neural networks

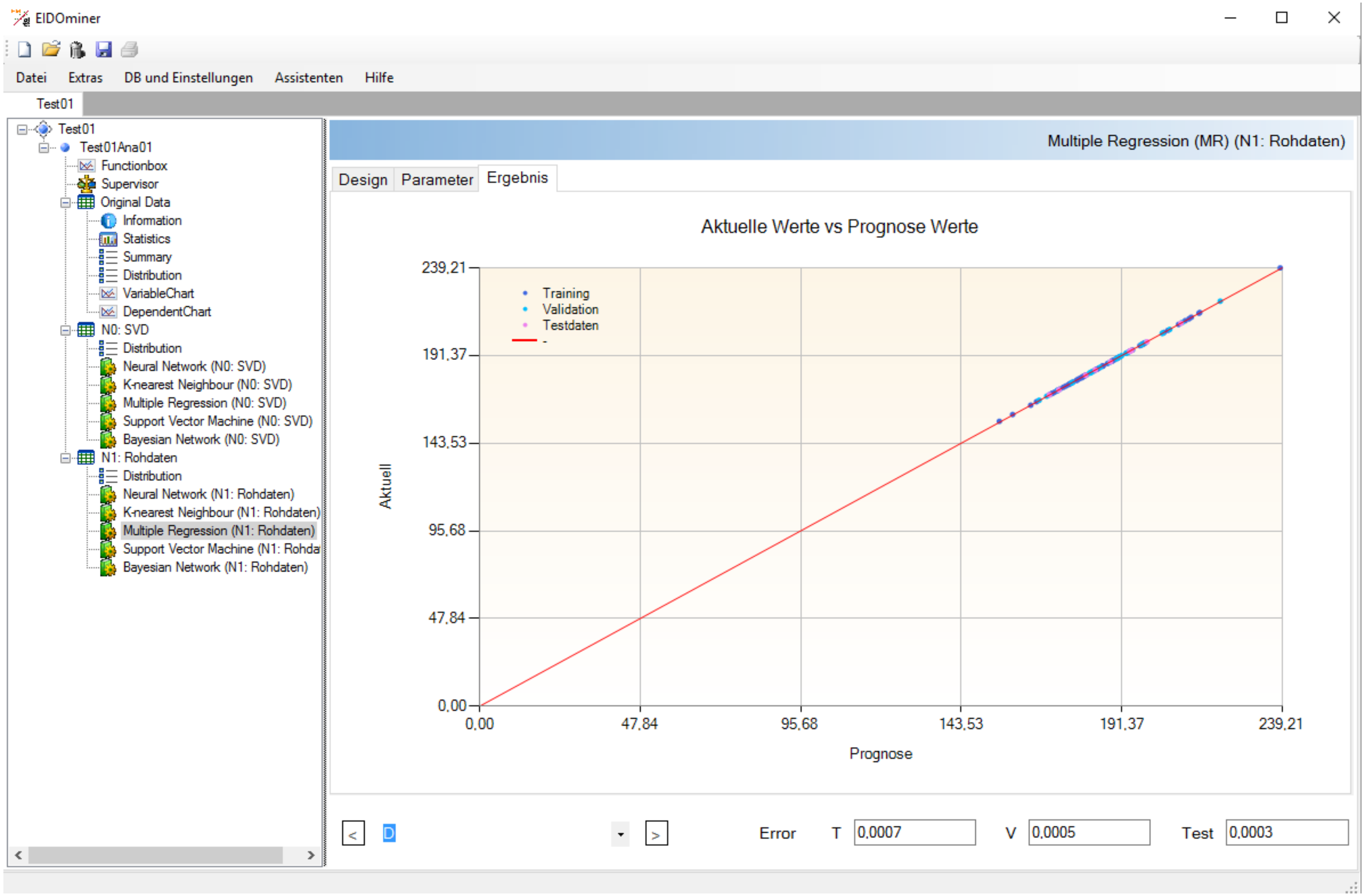


Predictive Analytics with EIDOminer

Result SVM



Result Multiple Regression



Hochschule Kempten

Universität Duisburg Essen – Offen im Denken

Multiple Regression (MR) (N1: Rohdaten)

Design | Parameter | Ergebnis

Variablen

Gleichungsvariablen

Result

Parameter	Coefficients	StandardError	T-Test	P-Value
I1	-3,4544	4,0598	-0,8509	0,3973
I2	0,4534	1,1593	0,3911	0,6967
I3	0,7684	0,8603	0,8931	0,3744
I4	7,2323	3,8604	1,8735	0,0645
I5	-1,5038	1,3307	-1,1301	0,2617
I6	0,4479	1,152	0,3888	0,6984
Constant	-32,2334	0,2634	-122,3661	0

Analysis of Variance

Information

R-squared = 99,992 percent
R-squared (adjusted for Df) = 99,992 percent
Standart error of Est. = 0,131
Mean absolute error = 0,082
The output shows the results of fitting a multiple regression model to describe the relationship between D and 7 independent variables.

The equation of the fitted model is

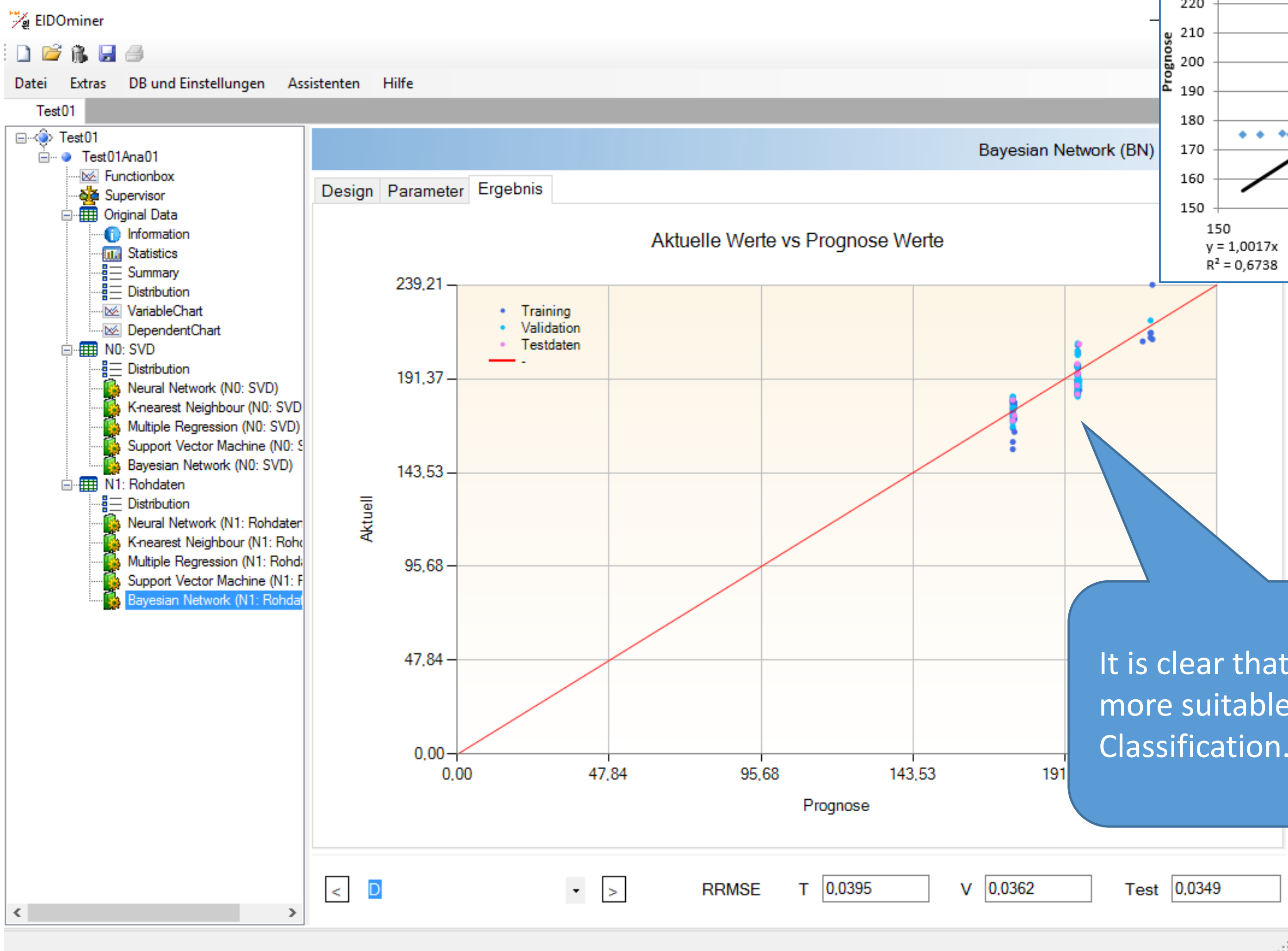
D = -3,45*(I1) + 0,45*(I2) + 0,77*(I3) + 7,23*(I4) + -1,50*(I5) + 0,45*(I6) + -32,23*(Constant)

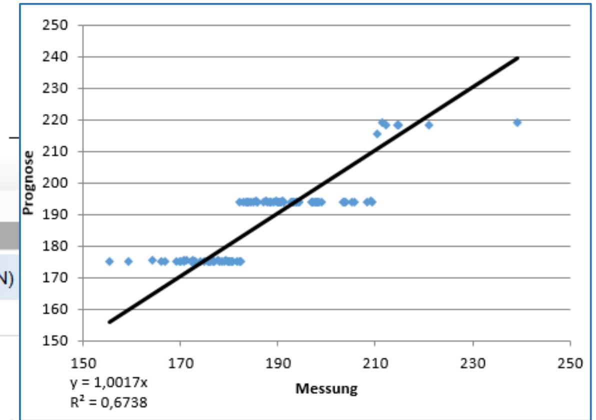
Start

However, the coefficients are not statistically significant.

Predictive Analytics with EIDOminer

Result Bayesian Network

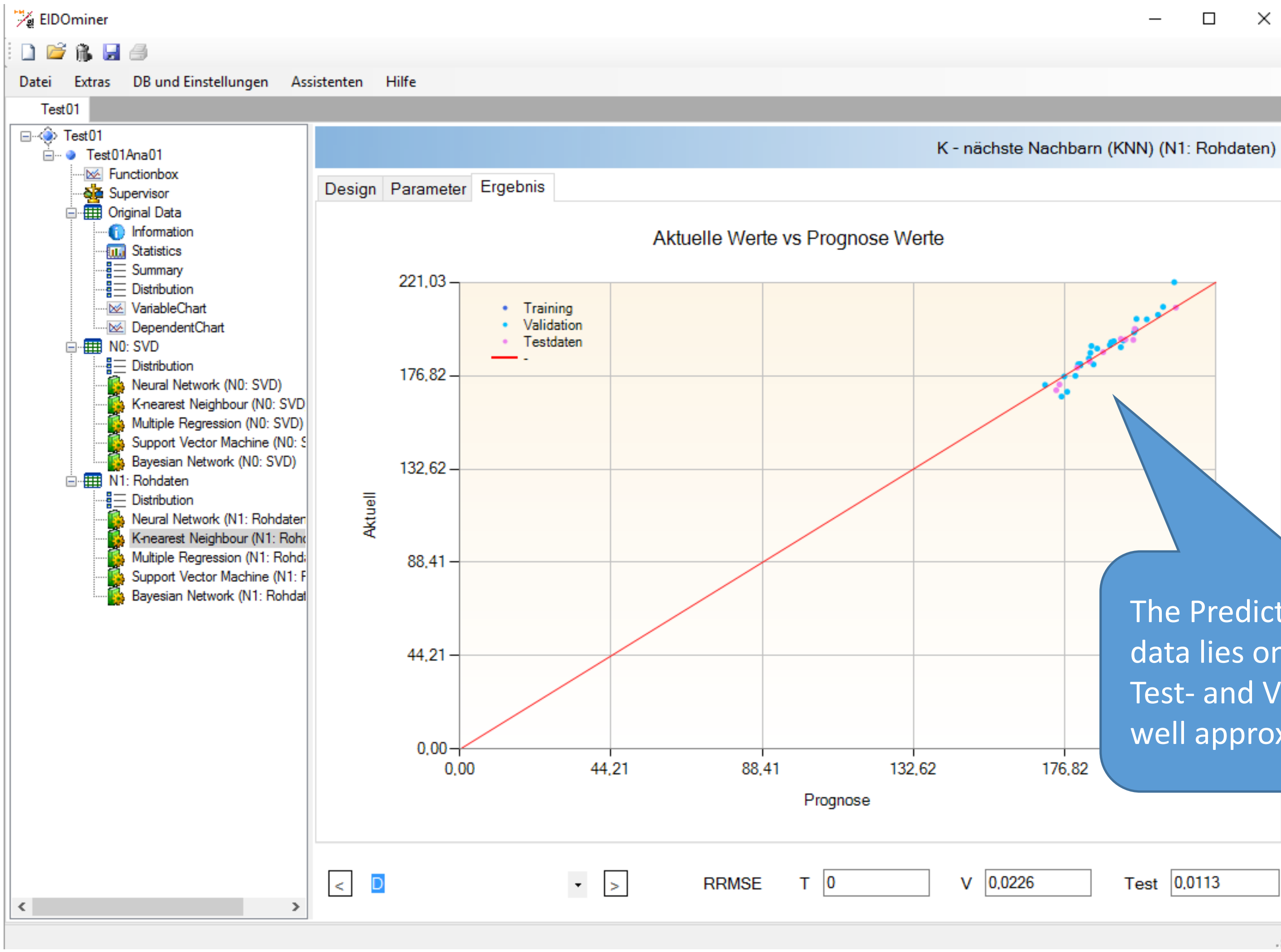




It is clear that this method is more suitable for Classification.

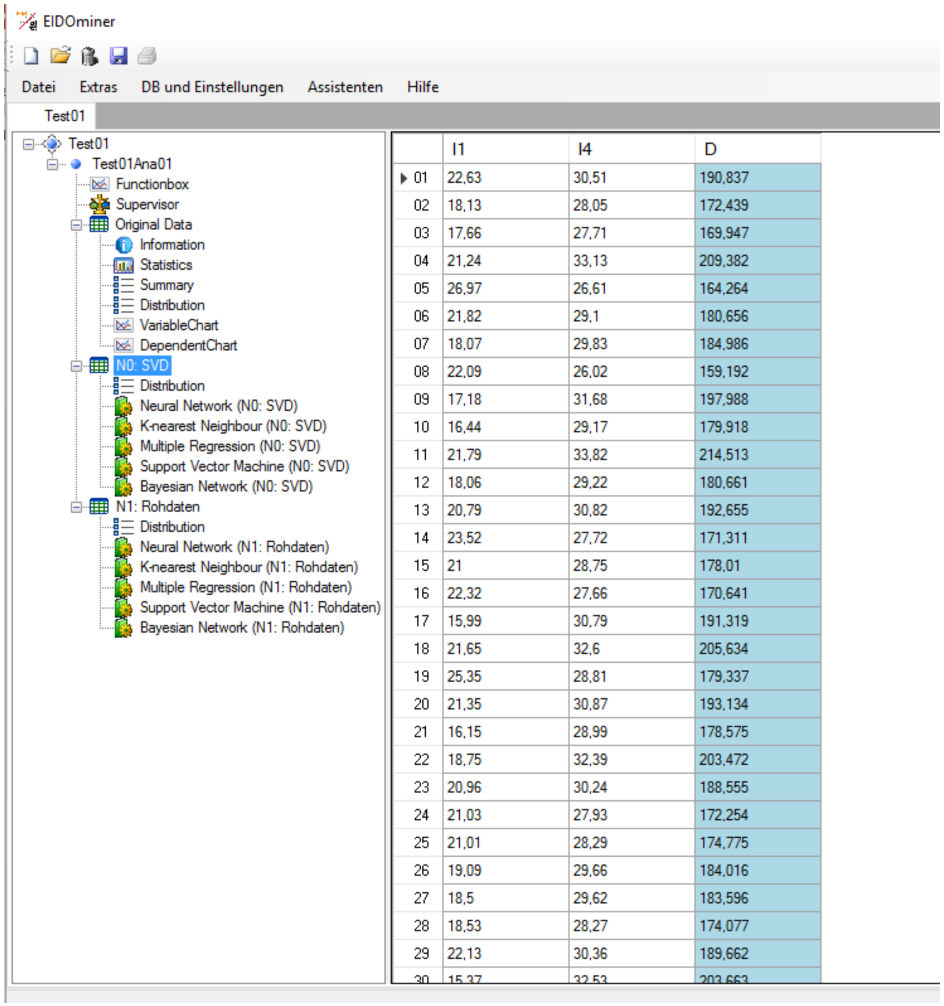
Predictive Analytics with EIDOminer

Result KNN



The Prediction of the Training data lies on the 45 ° axis. The Test- and Validation data are well approximated.

Dimension reduction)



	I1	I4	D
01	22,63	30,51	190,837
02	18,13	28,05	172,439
03	17,66	27,71	169,947
04	21,24	33,13	209,382
05	26,97	26,61	164,264
06	21,82	29,1	180,656
07	18,07	29,83	184,986
08	22,09	26,02	159,192
09	17,18	31,68	197,988
10	16,44	29,17	179,918
11	21,79	33,82	214,513
12	18,06	29,22	180,661
13	20,79	30,82	192,655
14	23,52	27,72	171,311
15	21	28,75	178,01
16	22,32	27,66	170,641
17	15,99	30,79	191,319
18	21,65	32,6	205,634
19	25,35	28,81	179,337
20	21,35	30,87	193,134
21	16,15	28,99	178,575
22	18,75	32,39	203,472
23	20,96	30,24	188,555
24	21,03	27,93	172,254
25	21,01	28,29	174,775
26	19,09	29,66	184,016
27	18,5	29,62	183,596
28	18,53	28,27	174,077
29	22,13	30,36	189,662
30	15,37	32,53	203,663

From the Singular Value Decomposition (SVD) in the preprocessing, all dependencies within the functionally independent variables are discovered. Only the variables I1 and I4 are relevant. The information of the other Independent variables is lost with this type of dimensional reduction.

Applying the Principal Component Analysis (**PCA**) to this data, the first two Principal Components **PC1** and **PC2** covers almost 100% of total variation. As with the SVD, the original data will be reduced to two columns. But with the SVD, it is the original columns **I1** and **I4**. In the case of the **PCA** both columns **PC1** and **PC2** knows about *all* columns of the original data.

Dimension reduction with the Principal Component Analysis - Principal Component Analysis (PCA)

Command Window

```
>> X

X =

   22.6300   67.8800  113.1700   30.5100   53.1500   95.8600
   18.1300   54.4000   90.6500   28.0500   46.1700   88.1200
   17.6600   52.9800   88.3100   27.7100   45.3800   87.0400
   16.8300   50.4800   84.1100   31.1000   47.9300   97.7000
   21.2400   63.7600  106.2200   33.1300   54.3800  104.0900
   26.9700   80.9200  134.8700   26.6100   53.5800   83.6300
   21.8200   65.4600  109.1000   29.1000   50.9200   91.4200
   18.0700   54.1900   90.3200   29.8300   47.9100   93.7200
   22.0900   66.2600  110.4500   26.0200   48.0800   81.7300
   17.1800   51.5200   85.9000   31.6800   48.8500   99.5500
   16.4400   49.3100   82.1900   29.1700   45.6100   91.6400
   21.7900   65.3700  108.9700   33.8200   55.6100  106.2600
   18.0600   54.1700   90.2700   29.2200   47.2700   91.8100
   20.7900   62.3700  103.9300   30.8200   51.5900   96.8300
   23.5200   70.5700  117.5900   27.7200   51.2200   87.0800
   21.0000   63.0200  105.0300   28.7500   49.7500   90.3100
   22.3200   66.9700  111.6400   27.6600   49.9800   86.9100
   15.9900   47.9500   79.9100   30.7900   46.7600   96.7200
   21.6500   64.9700  108.2600   32.6000   54.2500  102.4200
   25.3500   76.0300  126.7200   28.8100   54.1500   90.5200
   21.3500   64.0700  106.7600   30.8700   52.2400   96.9900
   16.1500   48.4400   80.7300   28.9900   45.1300   91.0800
   19.4900   58.4700   97.4700   30.2000   49.7100   94.8800
   18.7500   56.2400   93.7400   32.3900   51.1400  101.7700
   20.9600   62.8800  104.7700   30.2400   51.1800   95.0200
```

PCA

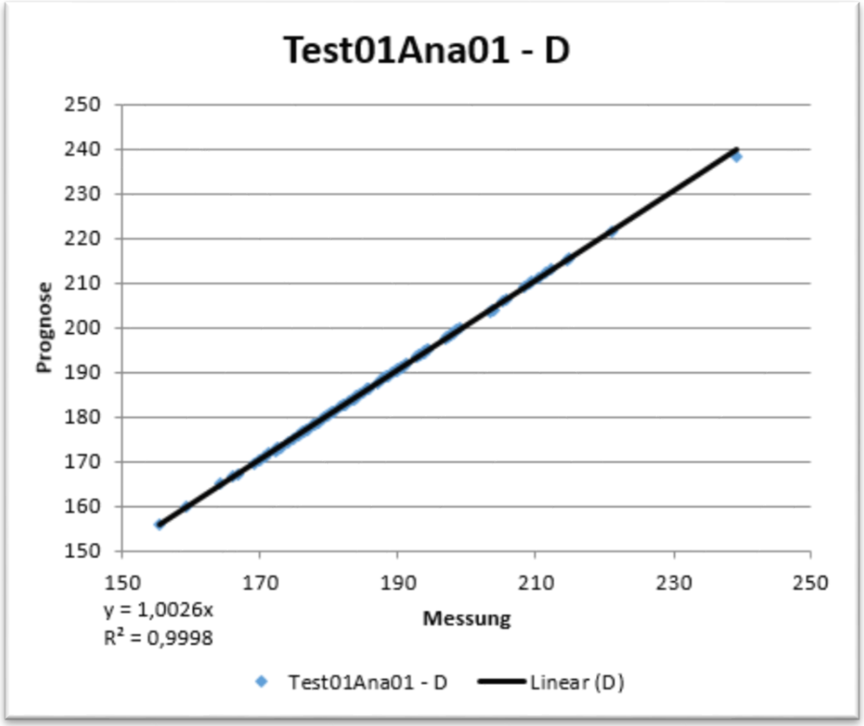
Command Window

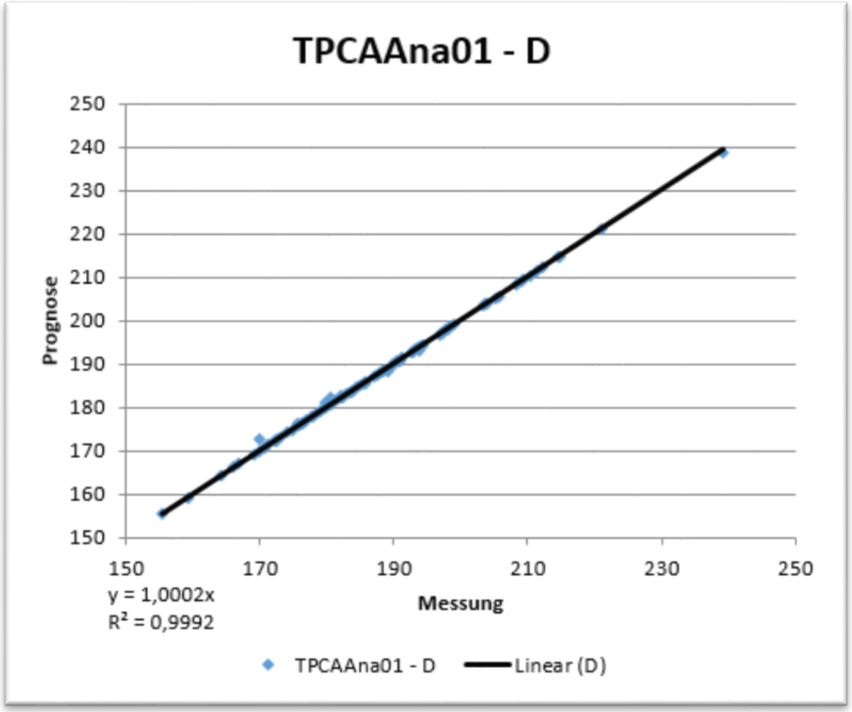
```
>> PC

PC =

 -136.1229  111.0979
 -109.1025  101.4186
 -106.2870  100.1181
 -101.3144  111.5847
 -127.8557  119.7564
 -162.0863   98.8245
 -131.2334  106.0141
 -108.7325  107.5484
 -132.7962   95.4528
 -103.4469  113.6977
  -98.9724  104.8354
 -131.1412  122.2781
 -108.6650  105.4442
 -125.0654  111.6681
 -141.3965  101.6984
 -126.3484  104.5822
 -134.2466  101.1975
  -96.2668  110.2836
 -130.2856  118.0314
 -152.3515  105.9540
 -128.4640  112.0023
  -97.2213  104.1409
 -117.3032  109.1984
 -112.8671  116.5515
 -126.0668  109.7259
```

Predictive Analytics with EIDOminer





RMSE = 0,49897209 (SVD)

Tools

Aktiviert	Trainiert	Name
☑	✓	Neural Network (N0: SVD)
☑	✓	K-nearest Neighbour (N0: SVD)
☑	✓	Multiple Regression (N0: SVD)
☑	✓	Support Vector Machine (N0: SVD)
☑	✓	Bayesian Network (N0: SVD)
☐	✓	Neural Network (N1: Rohdaten)
☐	✓	K-nearest Neighbour (N1: Rohdaten)
☐	✓	Multiple Regression (N1: Rohdaten)
☐	✓	Support Vector Machine (N1: Rohdaten)
☐	✓	Bayesian Network (N1: Rohdaten)

RMSE = 0,5318265 (Rohdaten)

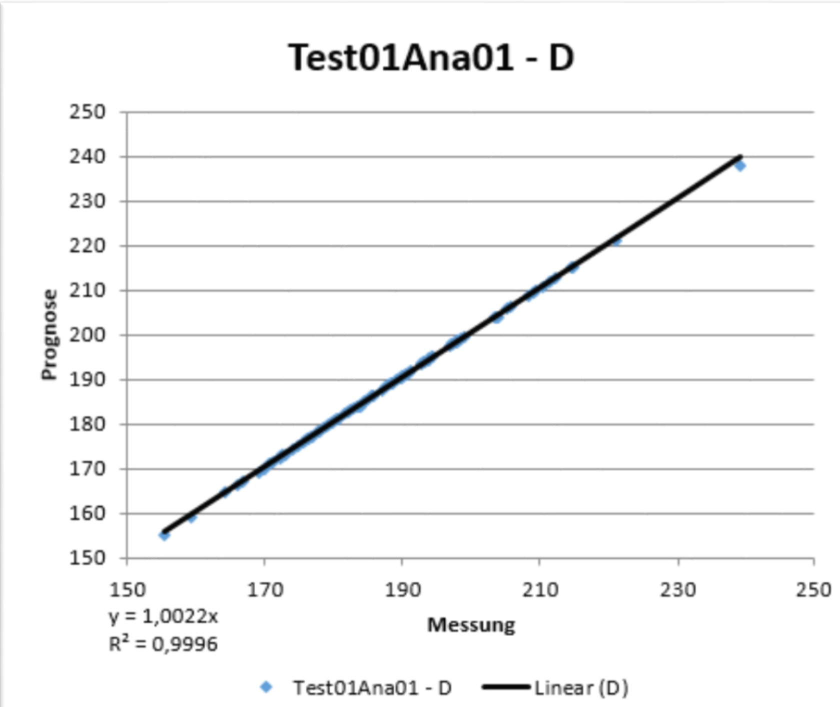
Tools

Aktiviert	Trainiert	Name
☐	✓	Neural Network (N0: SVD)
☐	✓	K-nearest Neighbour (N0: SVD)
☐	✓	Multiple Regression (N0: SVD)
☐	✓	Support Vector Machine (N0: SVD)
☐	✓	Bayesian Network (N0: SVD)
☑	✓	Neural Network (N1: Rohdaten)
☑	✓	K-nearest Neighbour (N1: Rohdaten)
☑	✓	Multiple Regression (N1: Rohdaten)
☑	✓	Support Vector Machine (N1: Rohdaten)
☑	✓	Bayesian Network (N1: Rohdaten)

Predictive Analytics with EIDOminer

Hochschule Kempten University of Applied Sciences

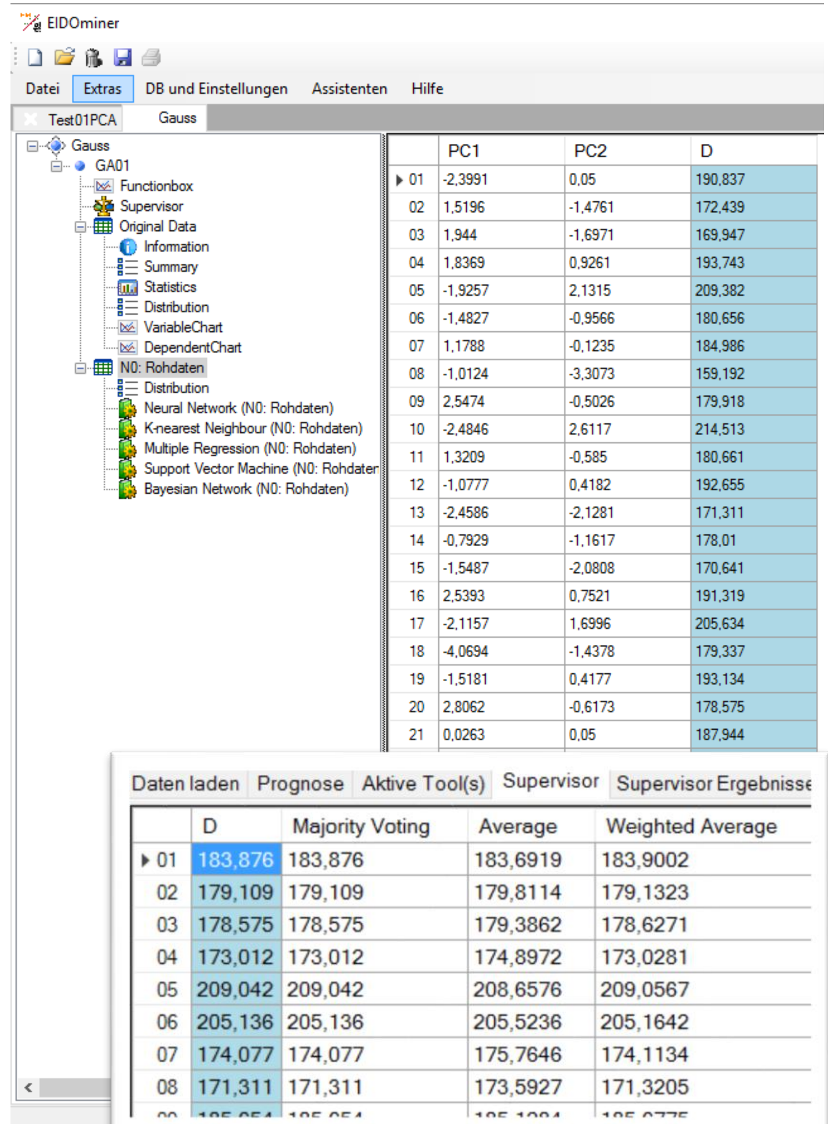
UNIVERSITÄT DUISBURG ESSEN Offen im Denken



RMSE = 0,3842543 (PCA)

Tools

Aktiviert	Trainiert	Name
☑	✓	Neural Network (N0: Rohdaten)
☑	✓	K-nearest Neighbour (N0: Rohdaten)
☑	✓	Multiple Regression (N0: Rohdaten)
☑	✓	Support Vector Machine (N0: Rohdaten)
☑	✓	Bayesian Network (N0: Rohdaten)



	PC1	PC2	D
01	-2,3991	0,05	190,837
02	1,5196	-1,4761	172,439
03	1,944	-1,6971	169,947
04	1,8369	0,9261	193,743
05	-1,9257	2,1315	209,382
06	-1,4827	-0,9566	180,656
07	1,1788	-0,1235	184,986
08	-1,0124	-3,3073	159,192
09	2,5474	-0,5026	179,918
10	-2,4846	2,6117	214,513
11	1,3209	-0,585	180,661
12	-1,0777	0,4182	192,655
13	-2,4586	-2,1281	171,311
14	-0,7929	-1,1617	178,01
15	-1,5487	-2,0808	170,641
16	2,5393	0,7521	191,319
17	-2,1157	1,6996	205,634
18	-4,0694	-1,4378	179,337
19	-1,5181	0,4177	193,134
20	2,8062	-0,6173	178,575
21	0,0263	0,05	187,944

	D	Majority Voting	Average	Weighted Average
01	183,876	183,876	183,6919	183,9002
02	179,109	179,109	179,8114	179,1323
03	178,575	178,575	179,3862	178,6271
04	173,012	173,012	174,8972	173,0281
05	209,042	209,042	208,6576	209,0567
06	205,136	205,136	205,5236	205,1642
07	174,077	174,077	175,7646	174,1134
08	171,311	171,311	173,5927	171,3205

Hochschule Kempten University of Applied Sciences

UNIVERSITÄT DUISBURG ESSEN Offen im Denken

Trained Prediction functions can give, with a sufficiently good datasets, *mathematically* justified good prediction for the functionally dependent variables on the basis of the functionally independent variables.

Every tool of Machine Learning has strengths and weaknesses, by intelligent fusion methods the predictions of the individual tools are strengthened and summarized into a common result.

If the functionally dependent Variables can not be predicted adequately from the functionally independent Variables even after a well-founded training phase (including data preparation), then, the database is not good.
As already mentioned, missing variables that have not been measured or are not known are the reasons for this. Furthermore, it should be checked whether the records can be classified by adding a new variable.

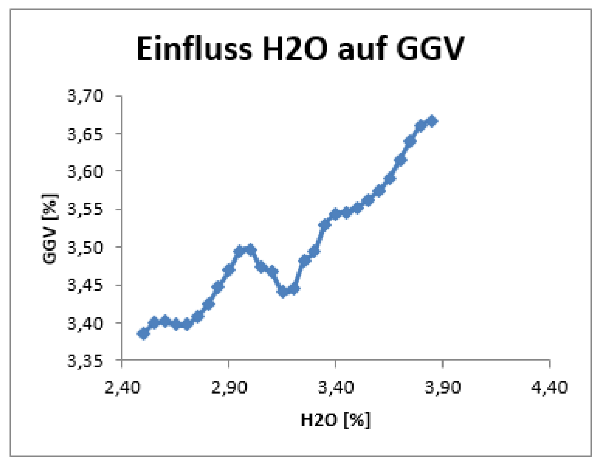
Examining the datasets and checking all process parameters that contain an influence on the dependent Variables give a deep insight into the investigated production process.

The console program *EIDOminerConsole* is the software to integrate trained prediction functions of the supervisor into external software. This allows the full range of functions of these programs (for example, MSExcel) to be used, for example, to create custom diagrams or perform parameter assignments.

History: Sand
Analysis: GGV
Range A7:E34

Predict Clear

Einwaage	H2O	C_Gehalt	Aktivton	SSG	GGV
150,50	2,50	2,36	7,60	11,37	3,39
150,50	2,55	2,36	7,60	11,37	3,40
150,50	2,60	2,36	7,60	11,37	3,40
150,50	2,65	2,36	7,60	11,37	3,40
150,50	2,70	2,36	7,60	11,37	3,40
150,50	2,75	2,36	7,60	11,37	3,41
150,50	2,80	2,36	7,60	11,37	3,42
150,50	2,85	2,36	7,60	11,37	3,45
150,50	2,90	2,36	7,60	11,37	3,47
150,50	2,95	2,36	7,60	11,37	3,49
150,50	3,00	2,36	7,60	11,37	3,50
150,50	3,05	2,36	7,60	11,37	3,47
150,50	3,10	2,36	7,60	11,37	3,47
150,50	3,15	2,36	7,60	11,37	3,44



In this example, the Water content is varied and the other functionally independent variables are kept constant (e.g. Median).

If the influence of a variable on the dependent Variable is not foundout, then it may be that this variable, in combination with *another* or even *several* independent variables, has a very strong *influence* on the Target variables (comparable to Cross-allergies).

Suppose there are 10 functional independent Variables and of each, only 10 variations are formed, then a Matrix with 10 Billion lines and ten columns is generated. The Prediction of this Matrix with the *EIDOminer* will be possible in the future ... today not yet.

In the new product *EIDOanalyzer* virtual Design of Experiment are integrated. With relatively few virtual “Trials", the influence of many factors on the target variables can be examined in order to recognize which factors are significant. The relationships between the "important factors" and the target variables can be determined with well-trained Prediction functions to determine optimal settings of the process parameters.

Hochschule Kempten University of Applied Sciences

UNIVERSITÄT DUISBURG ESSEN Offen im Denken

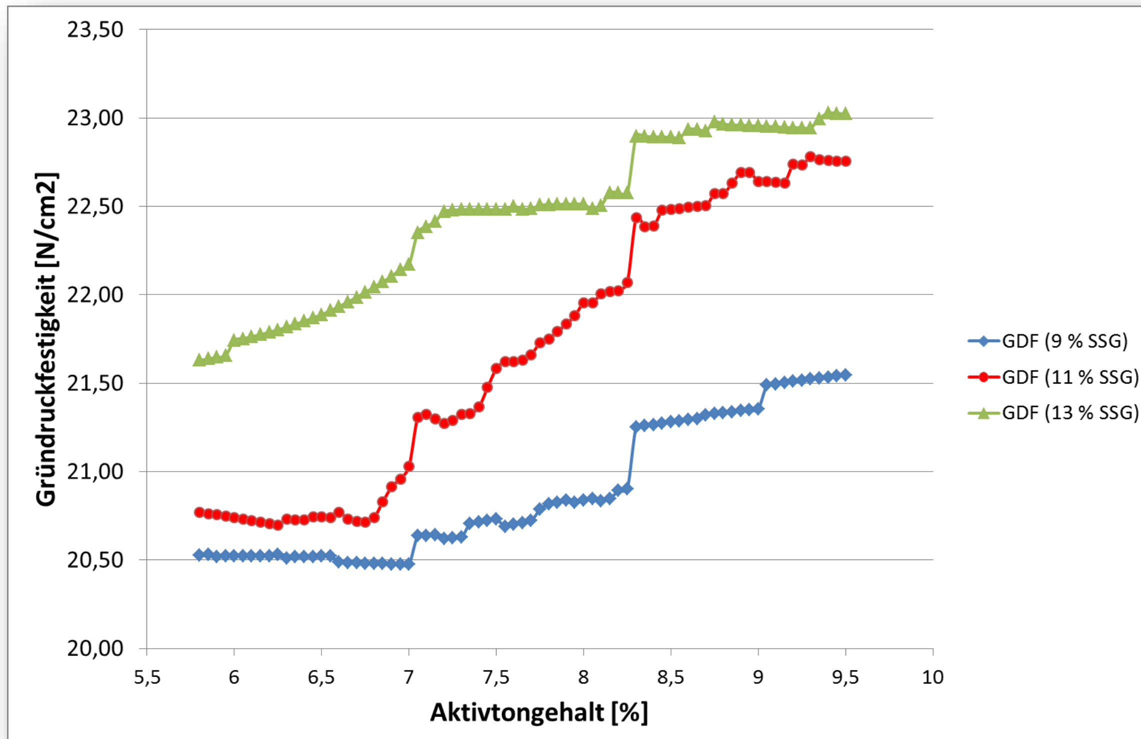
Process control via Parameter variation - Virtual DOE

History: Sand_oE
Analysis: DF
Range A7:D77

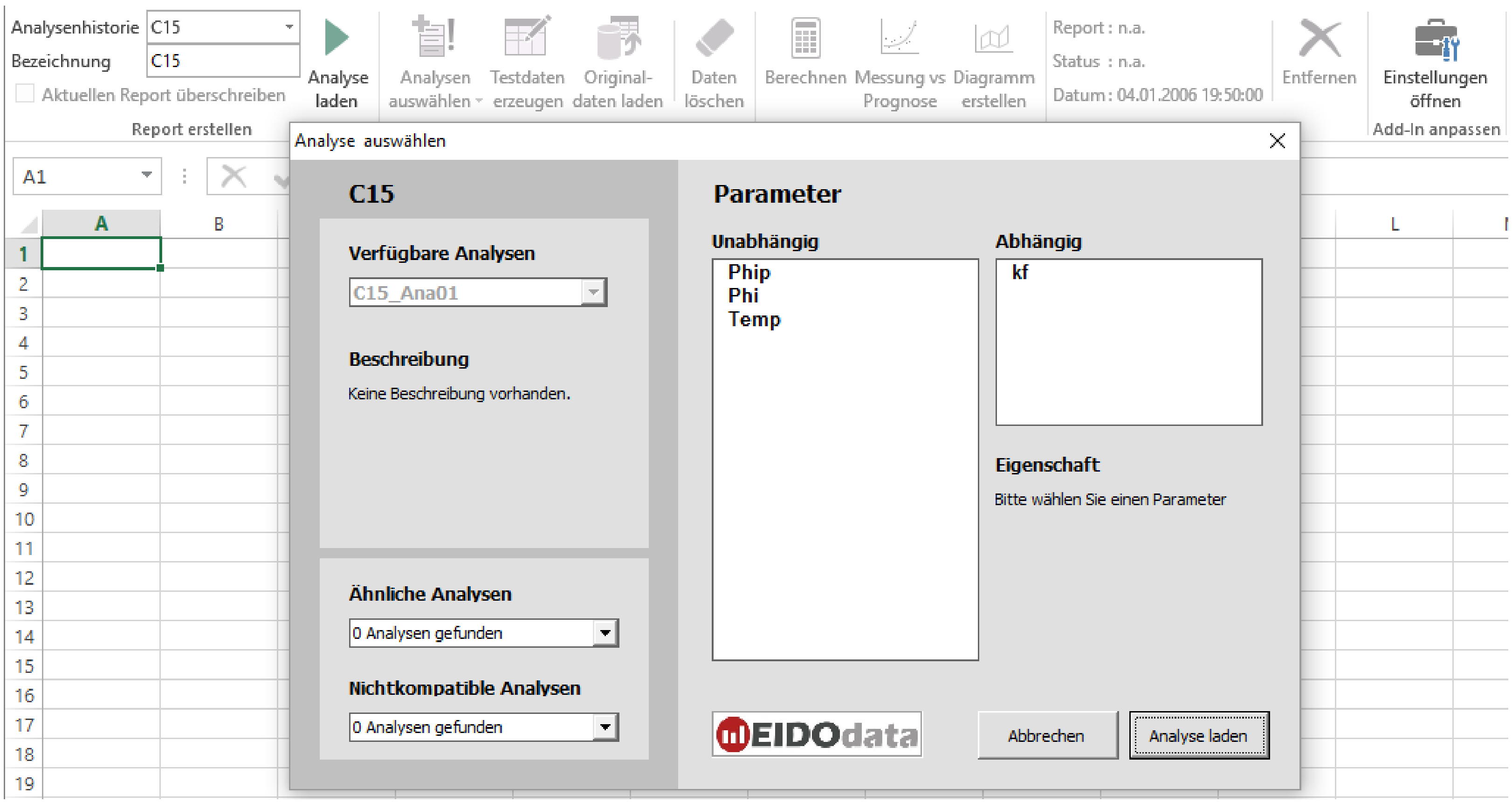
Predict

Clear

H2O	C_Gehalt	Aktivton	SSG	GDF
3,16	2,38	7,40	8,00	20,47
3,16	2,38	7,40	8,10	20,46
3,16	2,38	7,40	8,20	20,47
3,16	2,38	7,40	8,30	20,46
3,16	2,38	7,40	8,40	20,44
3,16	2,38	7,40	8,50	20,43
3,16	2,38	7,40	8,60	20,40
3,16	2,38	7,40	8,70	20,38
3,16	2,38	7,40	8,80	20,34
3,16	2,38	7,40	8,90	20,30
3,16	2,38	7,40	9,00	20,26
3,16	2,38	7,40	9,10	20,22
3,16	2,38	7,40	9,20	20,12
3,16	2,38	7,40	9,30	20,09
3,16	2,38	7,40	9,40	20,13
3,16	2,38	7,40	9,50	20,11
3,16	2,38	7,40	9,60	20,10
3,16	2,38	7,40	9,70	20,12
3,16	2,38	7,40	9,80	20,14
3,16	2,38	7,40	9,90	20,17
3,16	2,38	7,40	10,00	20,21
3,16	2,38	7,40	10,10	20,30
3,16	2,38	7,40	10,20	20,57
3,16	2,38	7,40	10,30	20,68
3,16	2,38	7,40	10,40	20,81



The *EIDOdata* software also contains the integrated Excel-Application *EIDOpredict*.

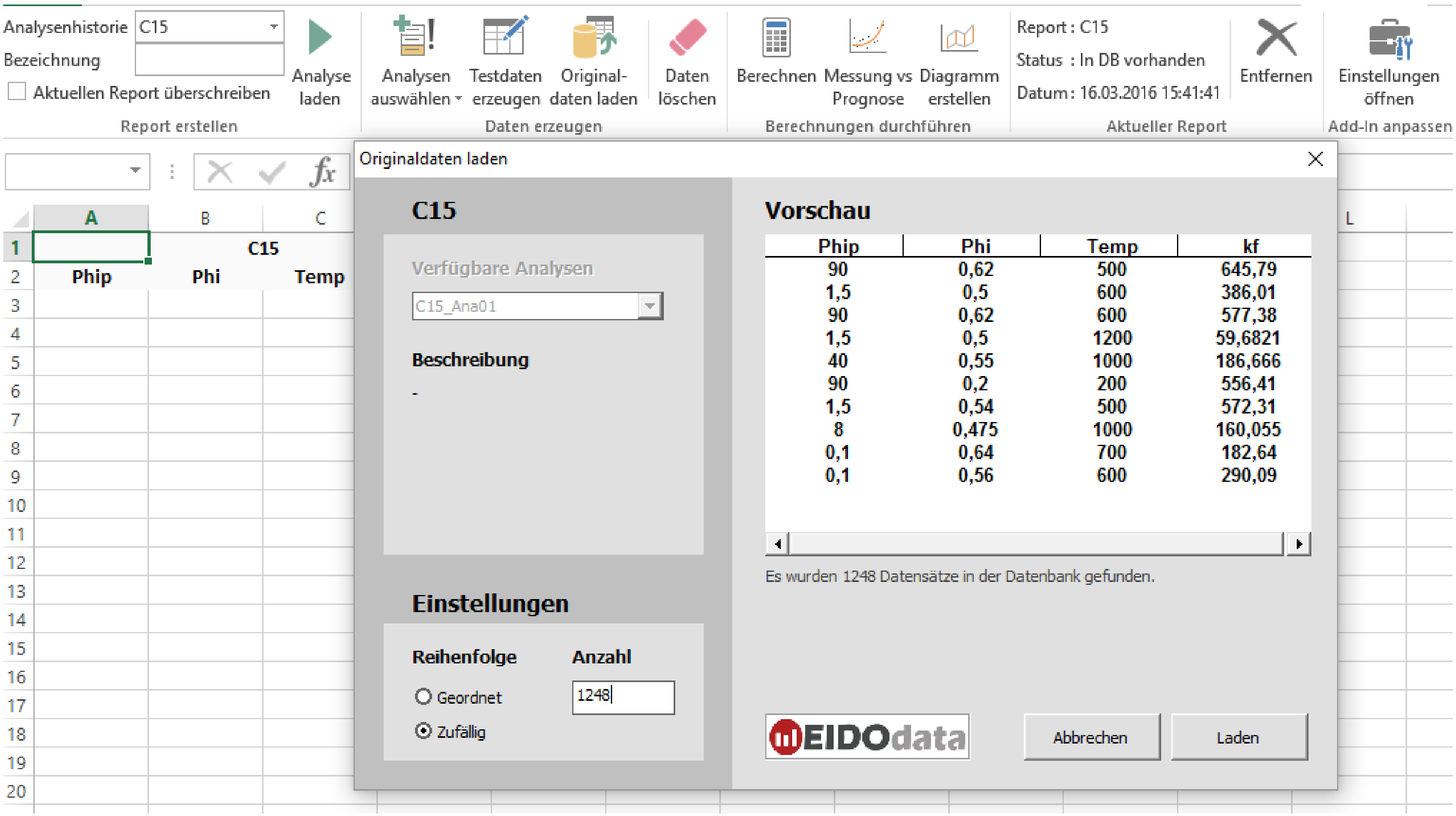


Predictive Analytics with EIDOminer

Hochschule Kempten

UNIVERSITÄT DUISBURG ESSEN *Offen im Denken*

Graphical Representations and Parameterizations are thus possible.



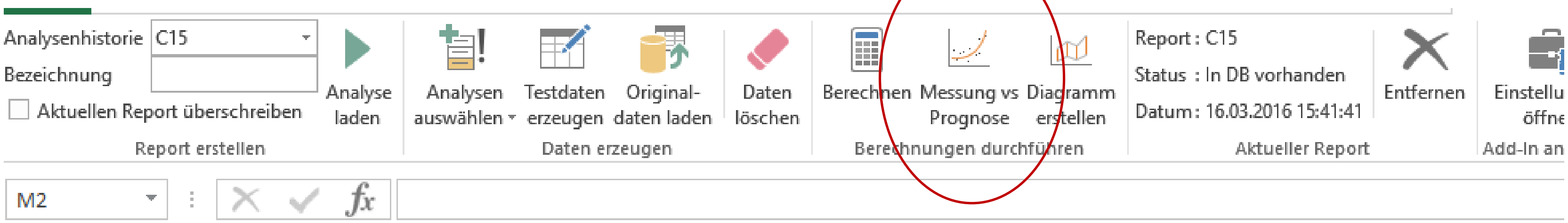
Phip	Phi	Temp	kf
90	0,62	500	645,79
1,5	0,5	600	386,01
90	0,62	600	577,38
1,5	0,5	1200	59,6821
40	0,55	1000	186,666
90	0,2	200	556,41
1,5	0,54	500	572,31
8	0,475	1000	160,055
0,1	0,64	700	182,64
0,1	0,56	600	290,09

Predictive Analytics with EIDOminer

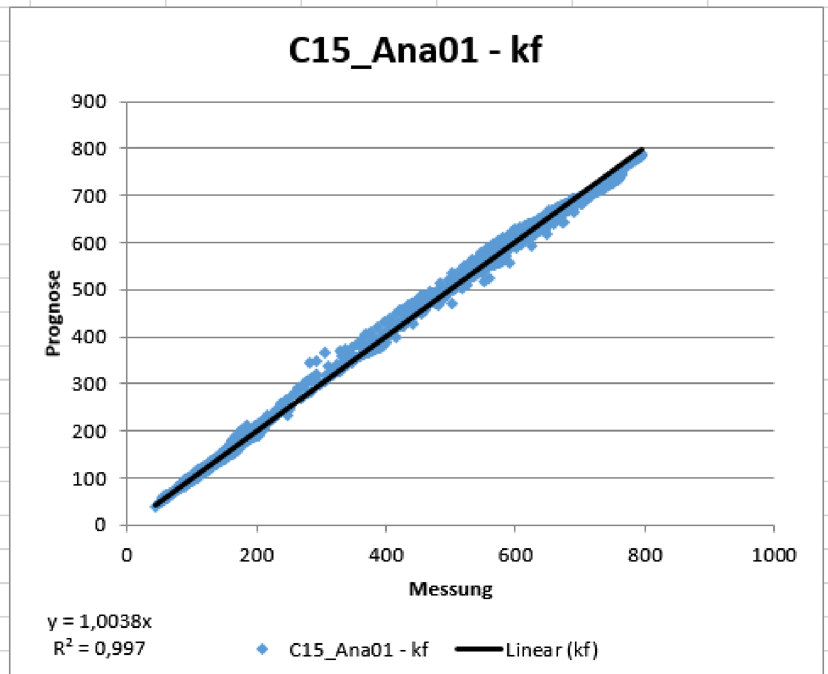
Hochschule Kempten

Universität Duisburg Essen – Offen im Denken

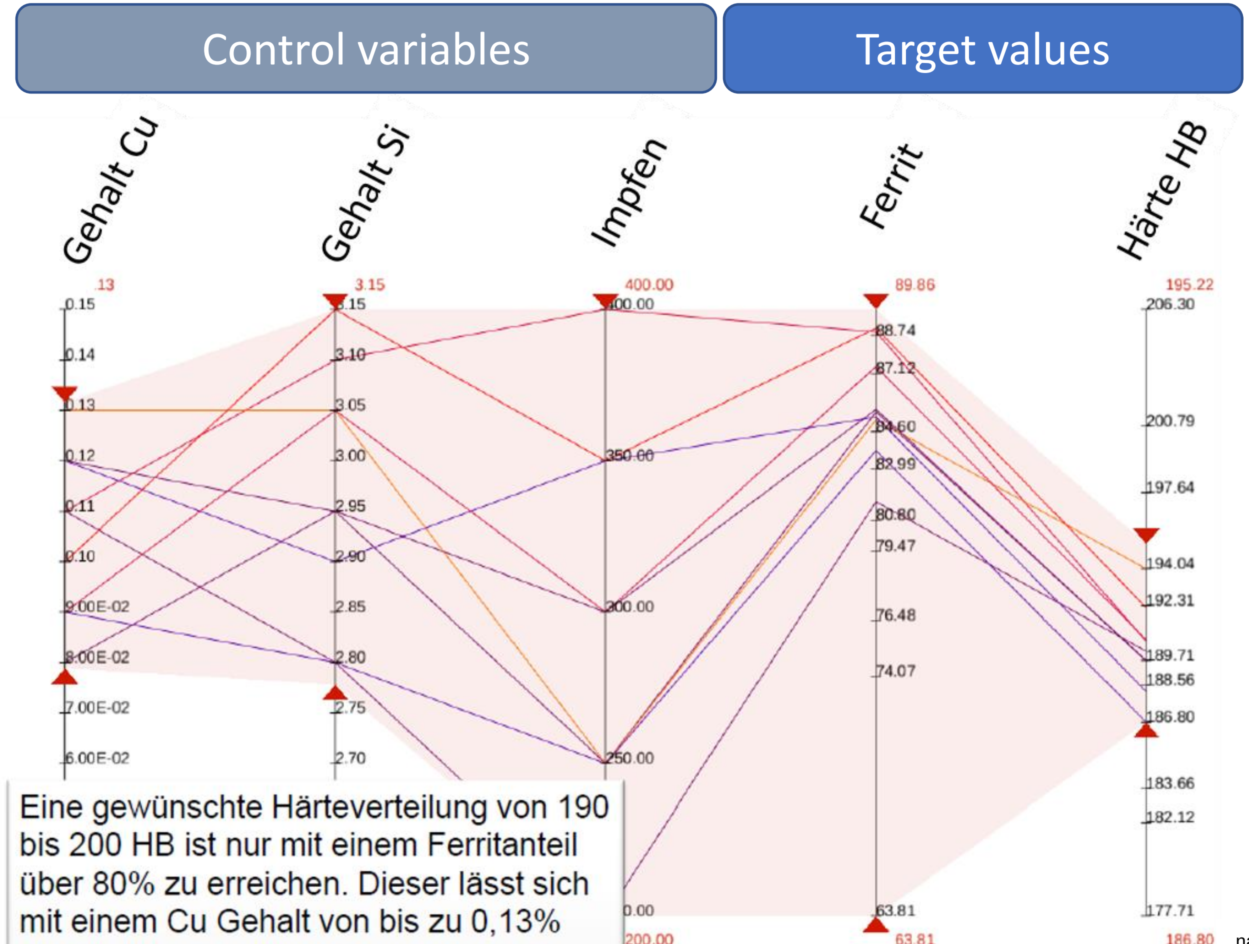
Prediction quality



	C15				C15_Ana01
	Phip	Phi	Temp	kf	kf
3	100	0,175	1100	136,983	133,319
4	1,5	0,66	100	686,02	674,969
5	0,1	0,46	700	176,51	198,376
6	90	0,7	200	672,4	678,26
7	1,5	0,04	100	338,77	342,106
8	1,5	0,34	100	632,96	635,789
9	40	0,15	1000	147,681	152,084
10	8	0,4	900	190,248	183,032
11	0,1	0,56	20	684,44	683,336
12	0,1	0,22	200	560,35	571,054
13	90	0,64	500	645,32	629,976
14	0,1	0,34	200	595,62	608,876
15	90	0,22	300	523,71	548,561
16	40	0,15	900	175,715	185,795
17	1,5	0,28	200	555,22	582,932
18	1,5	0,6	900	167,411	157,871
19	1,5	0,22	20	659,95	639,344
20	0,1	0,5	600	289,16	311,538
21	1,5	0,46	400	689,97	685,768



Parameter variation



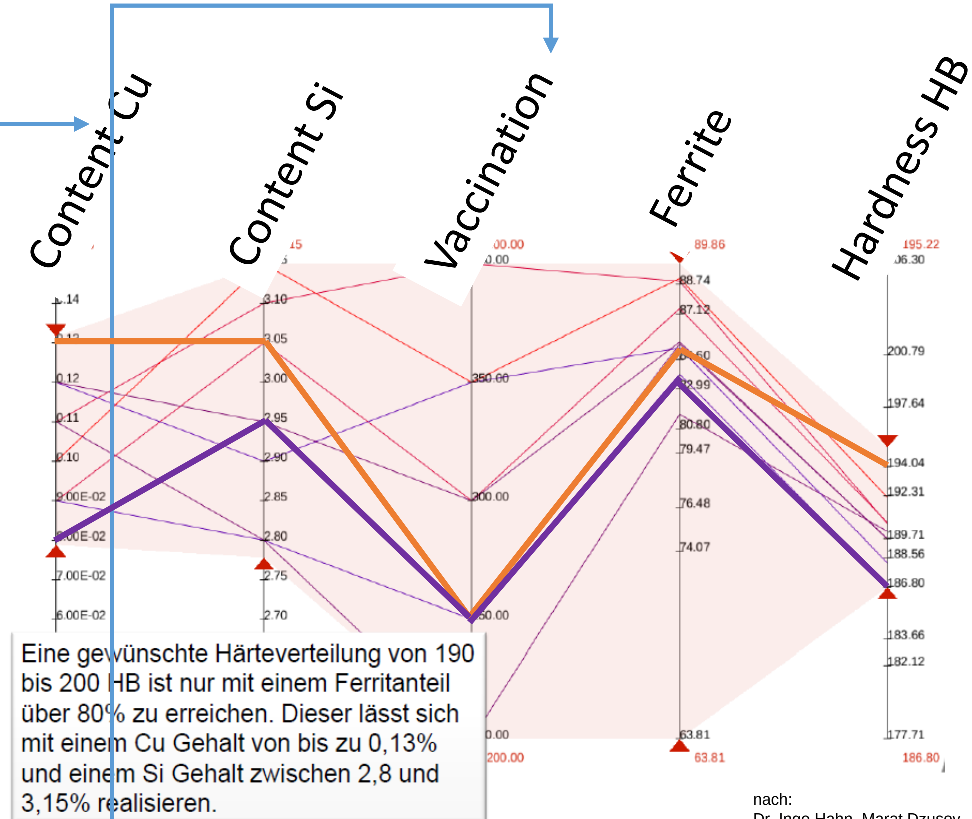
Eine gewünschte Härteverteilung von 190 bis 200 HB ist nur mit einem Ferritanteil über 80% zu erreichen. Dieser lässt sich mit einem Cu Gehalt von bis zu 0,13% und einem Si Gehalt zwischen 2,8 und 3,15% realisieren.

nach:
Dr. Ingo Hahn, Marat Dzusov
MAGMA Gießereitechnologie GmbH
8. MAGMA Eisenguss-Seminar 2014

Principal components and Sensitivity

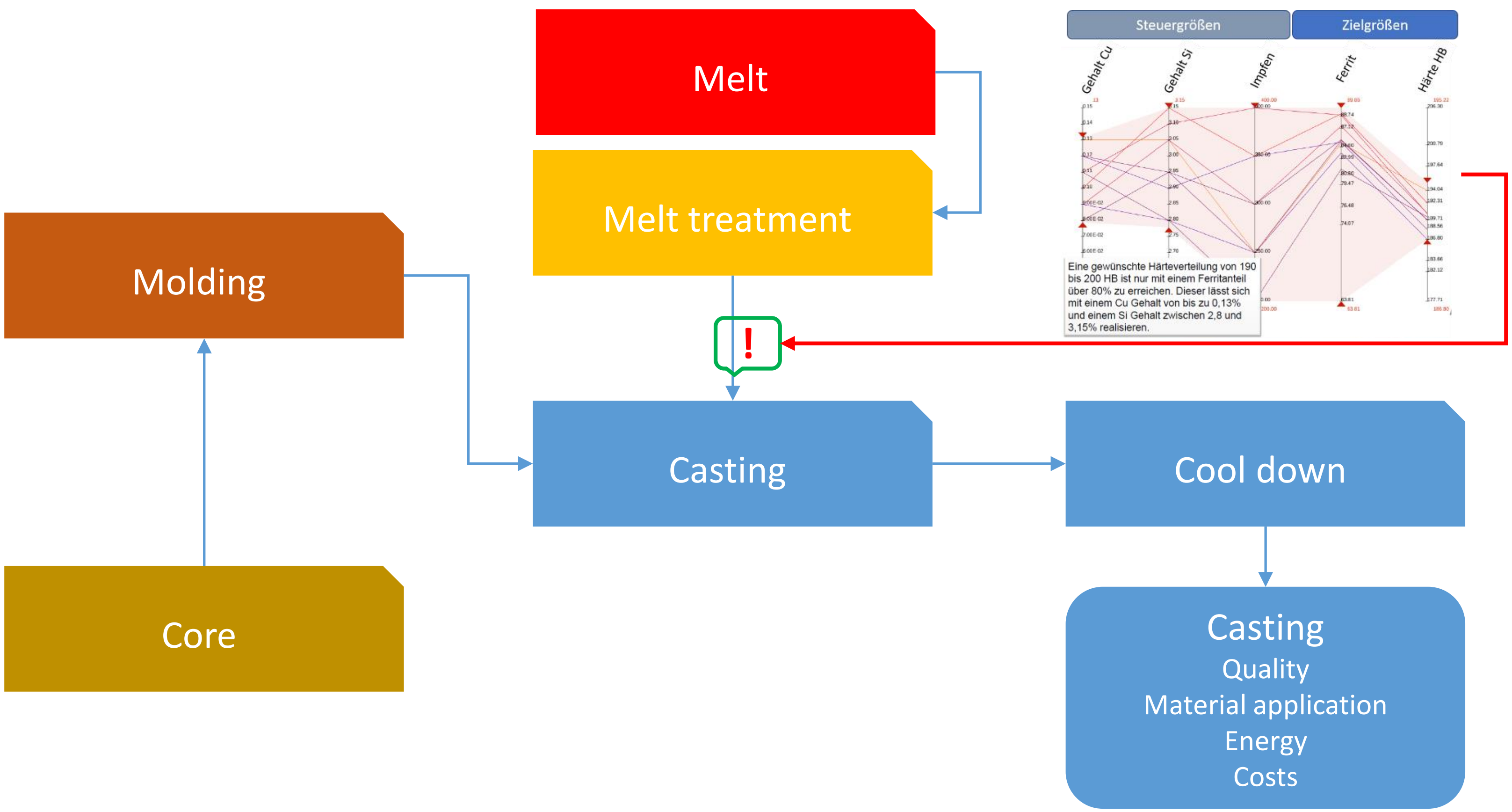
Most sensitive Parameter: Copper content

Principal Component analysis
Largest variation range:
Vaccination

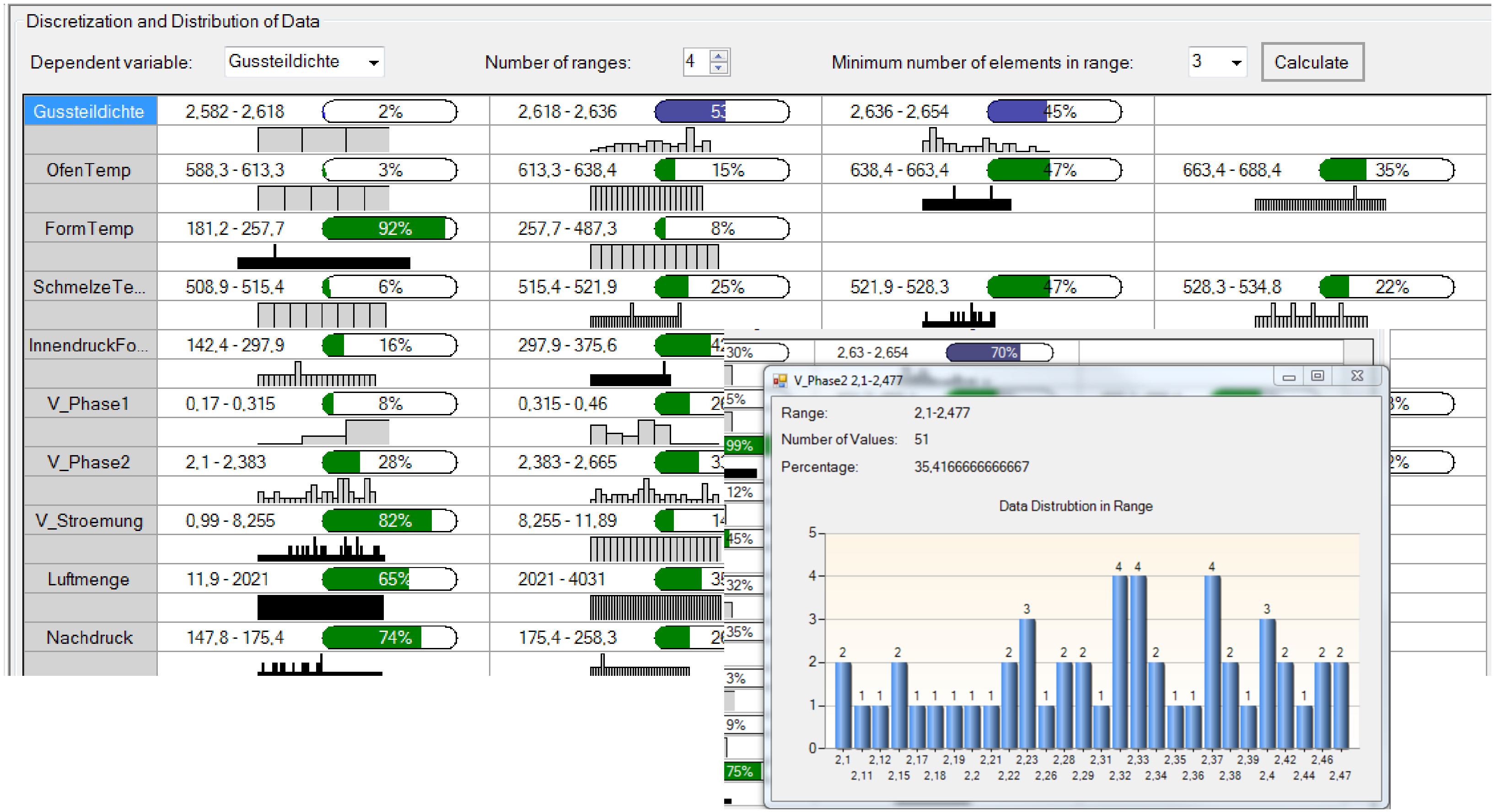


nach:
Dr. Ingo Hahn, Marat Dzusov
MAGMA Gießereitechnologie GmbH
8. MAGMA Eisenguss-Seminar 2014

Process control with Parameter variation

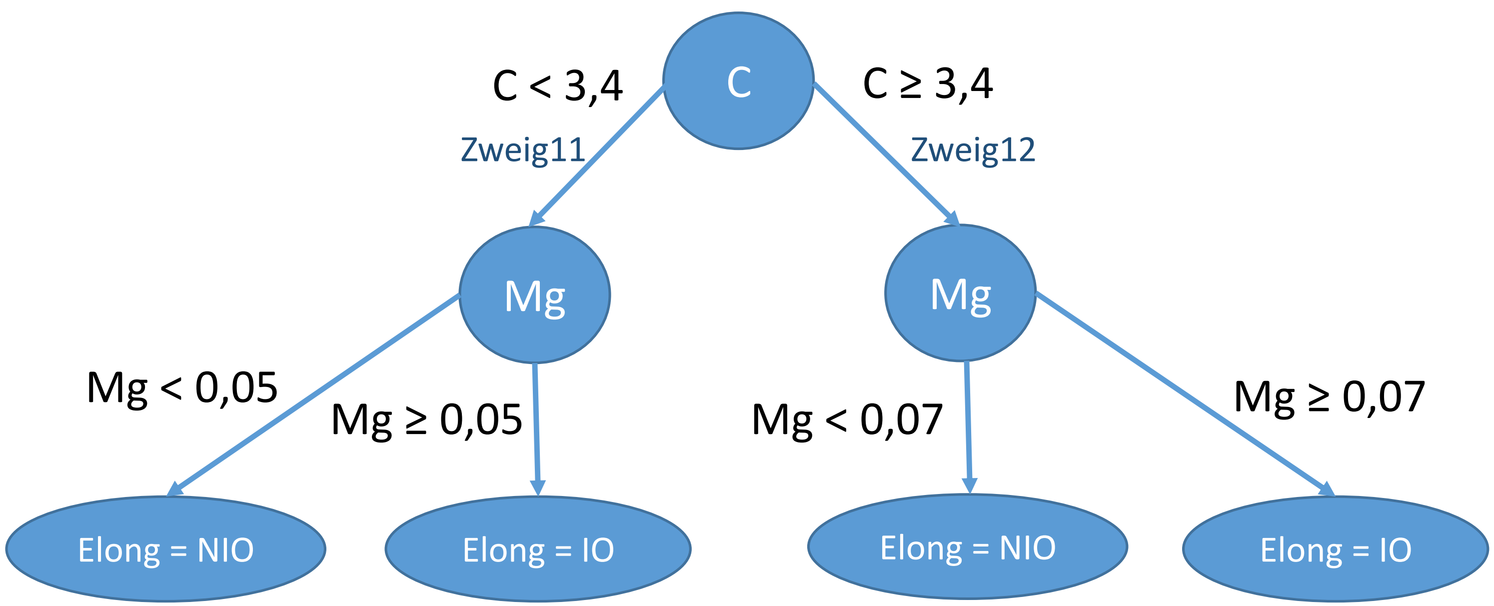


Statistical Analysis - EIDOdata



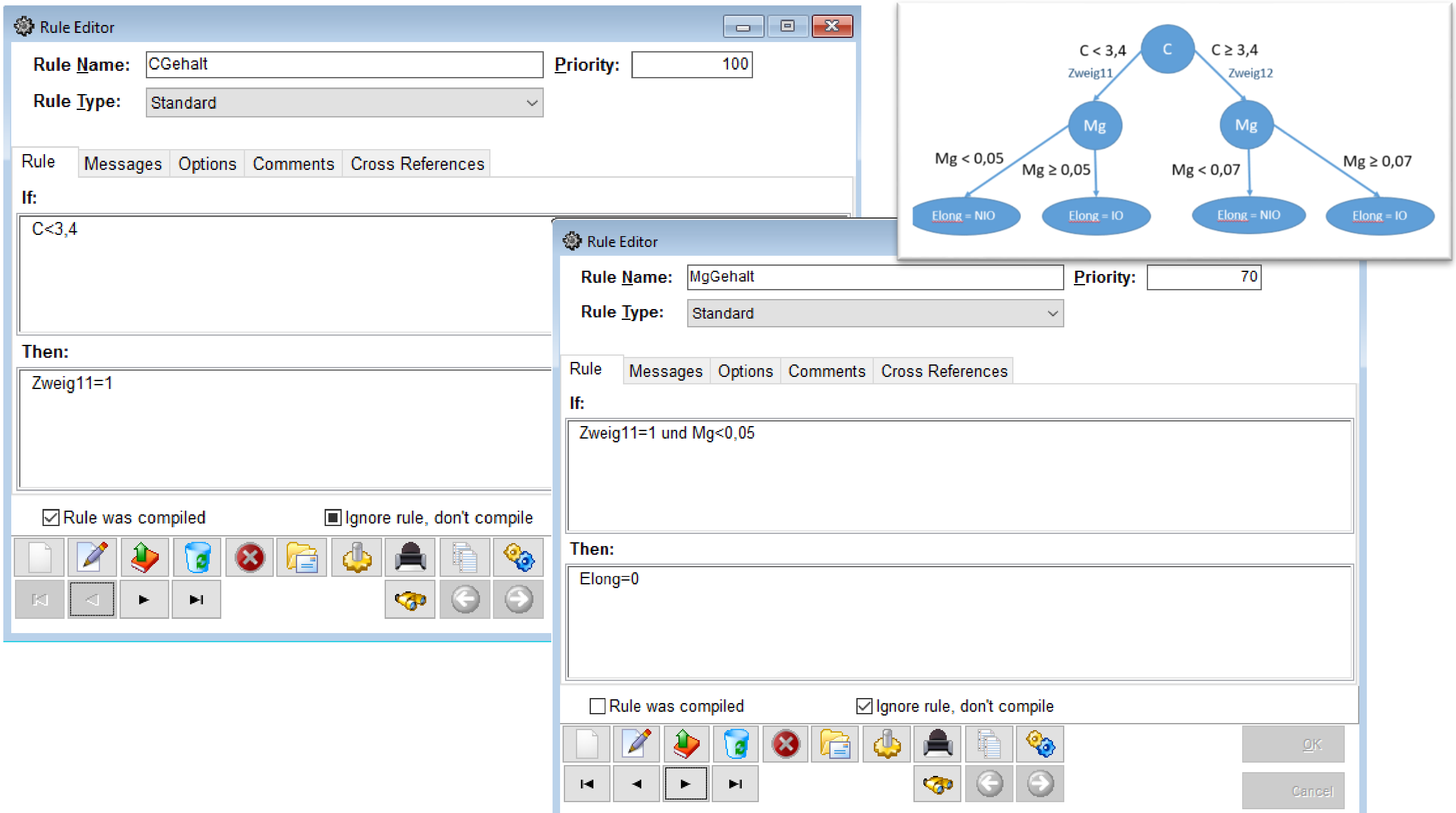
With the knowledge-based system *EIDOwiba,* one can create complex tree structures on the basis of if-then rules.

Here is a simple example of such a tree structure.



The Knowledge-base EIDOwiba

Such a tree structure is represented by the Rule-Editor as follows:

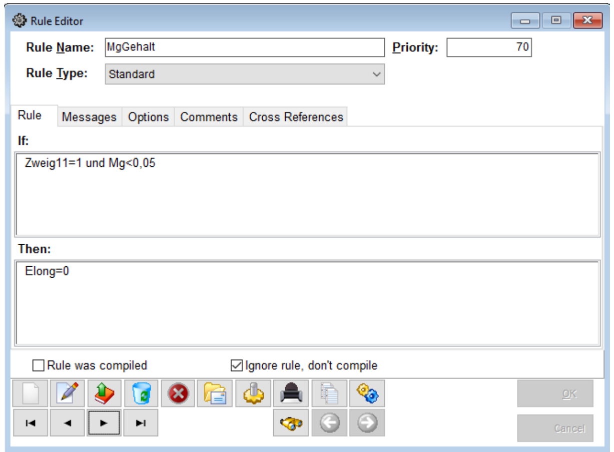


The Knowledge-base EIDOwiba

The knowledge base *EIDOwiba* in combination with well-trained Prediction functions of the *EIDOminer* can give in each production step optimized parameters to the process.

Knowledge is defined here as Experience and empirical Process knowledge in the form of rules and instructions of each Foundry as well as generally available knowledge, rules, correlations and formulas that are linked to the individual subprocesses.

The foundry module of the *EIDOwiba* is based on a basic knowledge of general rules of the casting technique, but must be adapted individually with regard to each sub-process.



Outlook

- The realization of Industry 4.0 methods has become a strategic competitive factor
- In production, the digital adaptive optimization of complex processes plays an important role
- Only collecting and selecting data is not a solution
- Generating Knowledge from data: The key to this is the intensive and adapted use of intelligent methods for data analysis
- Companies must be open minded to the problems of data acquisition, data analysis and intelligent use of predictions
- The Intelligence is still **AHEAD** of the computer: "Digitization" requires from the employees new skills and specialities