

# ***Machine Learning Basics***

***Seminarunterlage***

***Version: 1.01***



Dieses Dokument wird durch die ORDIX AG veröffentlicht.

Copyright ORDIX AG. Alle Rechte vorbehalten.

Alle Produkt- und Dienstleistungs-Bezeichnungen sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Firmen und beziehen sich auf Eintragungen in den USA oder USA-Warenzeichen.

Weitere Logos und Produkt- oder Handelsnamen sind eingetragene Warenzeichen oder Warenzeichen der jeweiligen Unternehmen.

Kein Teil dieser Dokumentation darf ohne vorherige schriftliche Genehmigung der ORDIX AG weitergegeben oder benutzt werden.

### **Adressen der ORDIX AG**

Die ORDIX AG besitzt folgende Geschäftsstellen

ORDIX AG  
Karl-Schurz-Straße 19a  
D-33100 Paderborn  
Tel.: (+49) 0 52 51 / 10 63 - 0  
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG  
An der alten Ziegelei 5  
D-48157 Münster  
Tel.: (+49) 02 51 / 9 24 35 – 00  
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG  
Welser Straße 9  
D-86368 Gersthofen  
Tel.: (+49) 08 21 / 507 492 – 0  
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG  
Kreuzberger Ring 13  
D-65205 Wiesbaden  
Tel.: (+49) 06 11 / 7 78 40 – 00  
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG  
Wikingerstraße 18-20  
D-51107 Köln  
Tel.: (+49) 02 21 / 8 70 61 – 0  
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG  
Gewerbegebiet Süd-West Park  
Südwestpark 67/2  
D-890449 Nürnberg  
Tel.: (+49) 0 52 51 / 10 63 - 0  
Fax.: (+49) 01 80 / 1 67 34 90

Internet: <http://www.ordix.de>

Email: [seminare@ordix.de](mailto:seminare@ordix.de)

## Inhaltsverzeichnis

<b>1</b>	<b>Einführung .....</b>	<b>5</b>
1.1	Einführung.....	6
1.2	Begriffsabgrenzung.....	7
<b>2</b>	<b>Lineare Regression .....</b>	<b>10</b>
2.1	Gliederung.....	11
2.2	Lineare Regression.....	13
2.3	Multivariate lineare Regression.....	17
2.4	Lineare Regression mit Basiserweiterung.....	21
2.5	Lineare Regression: Addendum.....	23
2.6	Statistische Grundlagen.....	24
2.7	Lagemaße.....	25
2.8	Streuemaße.....	27
2.9	Zusammenhangsmaße.....	29
2.10	Skalenniveaus.....	31
2.11	Kenngrößen der Regression.....	35
2.12	Kenngrößen der Regression: Addendum.....	40
2.13	Vor- und Nachteile von linearen Regressionsmodellen.....	41
<b>3</b>	<b>k-Nächste-Nachbarn .....</b>	<b>42</b>
3.1	Gliederung.....	43
3.2	k-Nächste-Nachbarn.....	45
3.3	Holdout-Validierung & Hyperparameter-Tuning.....	50
3.4	Holdout-Validierung.....	51
3.5	Hyperparameter-Tuning.....	52
3.6	AHoldout-Validierung & Hyperparameter-Tuning Addendum:.....	56
3.7	Kennzahlen der Klassifikation.....	58
3.8	Kennzahlen der Klassifikation: Addendum.....	63
3.9	Vor- und Nachteile von k-Nächsten Nachbarn.....	64
<b>4</b>	<b>Decision Trees .....</b>	<b>65</b>
4.1	Gliederung.....	66
4.2	Decision Trees.....	67
4.3	Decision Trees: Addendum.....	74
4.4	Explorative Datenanalyse.....	75
4.5	Explorative Datenanalyse: Addendum.....	83
4.6	Data Engineering: Variablenselektion & Data Cleansing.....	84
4.7	Data Engineering: Data Cleansing.....	86
4.8	Data Engineering: Variablenselektion.....	88
4.9	Data Engineering: Data Transformation.....	91
4.10	Overfitting & Underfitting.....	93
4.11	Decision Tree Pruning.....	97
4.12	Vor- und Nachteile von Decision Trees.....	100
<b>5</b>	<b>Ensemble-Modelle .....</b>	<b>101</b>
5.1	Gliederung.....	102
5.2	Ensemble-Modelle.....	103
5.3	Random Forests.....	107
5.4	AdaBoost.....	110
5.5	Vor- und Nachteile von Ensemble-Modellen.....	114
5.6	Data Science Pipeline.....	115
5.7	Komplexaufgabe.....	120
<b>6</b>	<b>Neuronale Netze .....</b>	<b>122</b>
6.1	Gliederung.....	123
6.2	Neuronale Netze.....	124
6.3	Backpropagation.....	129
6.4	Vertiefung neuronale Netze.....	132

6.5	Vor- und Nachteile von neuronalen Netzen .....	137
6.6	Data Engineering: Transformation & Feature Engineering.....	138
6.7	Data Engineering: Data Transformation .....	140
6.8	Data Engineering: Feature Engineering.....	145
6.9	Kreuzvalidierung .....	148
<b>7</b>	<b>Unsupervised Learning .....</b>	<b>153</b>
7.1	Gliederung.....	154
7.2	Machine Learning: Ein Überblick .....	155
7.3	Clustering .....	159
7.4	K-means.....	161
<b>8</b>	<b>Schlussteil.....</b>	<b>166</b>
8.1	Gliederung.....	167
8.2	Modellüberblick .....	168
8.3	Technologie-Stack .....	170
8.4	Best Practices .....	172
8.5	Feedbackrunde .....	174