

Inhaltsverzeichnis

Zusammenfassung	2
Inhaltsverzeichnis	3
Abbildungsverzeichnis	6
Einführung in Data Mining.	7
Definition.	7
Einleitung.	7
Einheitliche Begriffsterminologie	8
Abgrenzung zu anderen Anwendungen	8
Anwendungsgebiete des Data Mining.	8
Überwachtes und unüberwachtes Lernen.	10
Methoden.	10
Klassifikation.	10
Clusteranalyse	11
Bayes-Ansatz bei Clusteranalyse.	11
Problemstellung.	13
Datenverarbeitung / -erfassung bei HT.	13
Software. Informationssystem Aristoteles®	13
Formularen.	15
Datenorganisation im Modul DIVA.	16
Datenverfremdung.	17
Zielsetzung.	17
Interessantheit	18
Anwenderbezogenheit	19
Anzahl den Aussagen.	20
Anomalien	20
Interpretierfähigkeit/Verständlichkeit	20
Kosten	21
Abdeckung (Coverage)	21
Spezifische Zielsetzungen bei Hochtief.	21
Implementierung (Ideen).	23
KDD Prozeß.	23
1 Geschäftsverständnis	23
2. Datenverständnis	25
3. Datenaufbereitung	26
4. Data Mining	27
5. Auswertung	28
6. Implementierung	29
Ablauf von KDD Prozesses bei Hochtief.	30
Datenfluß.	31
Problemlösung (Ablauf).	33
Ablauf.	33
Auswahl von Diplomzielen.	33
Klassifikation	33

Prognose	33
Suche nach Assoziationsregeln	34
Segmentierung	34
Auffinden informativer Fälle	34
Datenqualität.	34
Datenaufbereitung. Vorverarbeitung Operationen.	35
Selektion	35
Bereinigung	35
Gruppierung	36
Konstruktion neuer Information	37
Visualisierung	37
Definition von Zielklassen.	38
Ergebnis in absoluten Werten in DM	38
Ergebnis der Investition in prozentualen Werten	38
Gemischte Form von Klassen	39
Funktionale Form. Ergebnis als $F(\%,DM)$	39
Charakterisierung der Top- Flop- Projekte	40
Praxis, Testphase, Ablauf, Testbeschreibung.	41
Vorgegebene Zielklassen.	41
Von Hochtief vorgeschlagene Zielklassen.	41
Mit oder ohne genau 0.	42
Mit einer Zielklasse symmetrisch gegen 0.	42
Mit einer Zielklasse symmetrisch gegen Maximum.	43
Zielklassen in der gemischten Form.	44
Zielklassen als Funktion.	45
Zielklassen per Autoclass definiert.	47
Autoclass	47
Anzahl von Datensätzen variieren.	49
Klassifikation mit Support Vector Machine (SVM).	50
Gewinn in % als Zielattribut	50
Gewinn in DM als Zielattribut	51
SVM –C4.5 Vergleich	52
Analyse mit kNN und BNGE.	52
Analyse von Baustellen gleiche Art.	53
Bilden von Zielklassen „Richtig“ und „Falsch“.	54
Untergruppenanalyse mit MIDOS.	54
Zielattribut in Prozenten	55
Gemischte Form von Zielklasse.	58
Funktionale Form von Zielklassen.	59
Fazit	62
Top - Flop 10 Analyse.	63
Flop 10 %	63
Top & Flop 10 Baustellen	64
Eingesetzte Verfahren	65
C4.5.	65
Beschreibung, Einführung	65
Support Vector Machines (SVM).	65
Vorteile gegen C4.5	66
kNN.	67
BNGE.	67
Autoclass	67

Midos	67
Kepler	68
Beschreibung	68
Plattformen	69
Datenzugriff	69
Integration von mehreren Data Mining Verfahren.	69
Verwalten von Analyseergebnissen und Zwischenschritten.	70
Automatische Umwandlung von Formaten	71
Visualisierung	71
Mögliche Weiterentwicklung	73
Einführung von Data Mining in Unternehmen.	73
Data Warehouse.	73
Erfassung von weiteren Informationen.	74
Informationsimport aus anderen Datenquellen.	74
Manuelle Erfassung	75
Ergebnisse.	76
Richtlinien in der Entwicklung von Data Mining Tools.	76
Kombination verschiedener Methoden	76
Datenvorverarbeitung	76
Anwendungsintegration	76
Verifizierung der Ergebnisse	76
Bauprofil von HT.	76
Abhängigkeit zwischen den Datensätzen.	77
Zeitliche Veränderungen.	77
Datensätze aus derselbe Verteilung.	77
Mangel an Information. (Attributen).	78
Fazit	78
Literaturverzeichnis.	79
Anhang.	82
A. Formularen.	82
B. Struktur von SProj.	83
C. Katalog ART.	85
D. Anzahl von Baustellen mit 1%,2%,3% u.s.w. Gewinn.	86
E. Auswahl von Parameter mit SVM.	88
F. Verteilung von dem Attribut Architekt.	90
G. Gruppierung von dem Attribut Bauwerksart	91
H. Entscheidungsbaum von C4.5 (Teil)	92
I. Anwendung auf Daten von HOCHTIEF	94

Abbildungsverzeichnis

ABBILDUNG . Zuordnung von Data Mining als Wissenschaft.	8
ABBILDUNG. Aristoteles()	14
ABBILDUNG. Datenfluß Aristoteles	15
ABBILDUNG. Datenorganisation im Modul DIVA	16
ABBILDUNG. KDD- Prozeß als Zyklische Entwicklung	23
ABBILDUNG. Data Mining Prozeß als Zyklische Entwicklung	30
ABBILDUNG. Datenrepräsentation	31
ABBILDUNG. Prognose	33
ABBILDUNG. Zielklassen als Kombination aus zwei Variablen.	39
ABBILDUNG. Verteilung von Beispielen in 9 Klassen.	39
ABBILDUNG. Verteilung von Beispielen in 3 Klassen durch eine lineare Funktion.	40
TABELLE: Die vorgegebenen Klassen mit der Grenze am 0.	41
TABELLE: Die 4 bis 5 Klassen.	42
TABELLE: Untersuchung von der Grenze bei 0.	42
TABELLE: Eine Zielklasse symmetrisch gegen den 0.	43
ABBILDUNG: Zielklassendefinition bei der Glockenkurve.	43
TABELLE: Eine Zielklasse symmetrisch gegen den Maximum.	43
TABELLE: Verteilung von Beispielen in 9 Klassen.	44
TABELLE: Zielklassenmatrix mit 9 Klassen.	44
TABELLE: Verteilung von Beispielen in 9 Klassen.	44
TABELLE: Zielklassenmatrix mit 9 Klassen.	45
ABBILDUNG: Zielklassen durch eine Funktion getrennt. Eine Klasse gegen 0	45
TABELLE: Zielklassen durch eine Funktion getrennt.	46
ABBILDUNG: Zielklassen durch eine Funktion getrennt. Grenze durch 0.	46
TABELLE: Zielklassen durch eine Funktion getrennt.	47
ABBILDUNG: Autoclass. Verteilung von 4 Zielklassen.	48
ABBILDUNG: Autoclass. Verteilung von 8 Zielklassen	48
TABELLE: Zielklassen per AutoClass definiert.	49
ABBILDUNG: Einfluß von der Datenmenge.	49
TABELLE: SVM. Tests mit dem Ergebnis in %.	51
TABELLE: SVM. Tests mit dem Ergebnis in %.	51
TABELLE: SVM und C4.5 Vergleich.	52
TABELLE: Analyse mit BNGE.	53
TABELLE: Analyse von Baustellen gleiche Art.	53
ABBILDUNG: Zielklassen mit eine Grenze am 0.	55
ABBILDUNG: Zielklassen symmetrisch von 0.	56
ABBILDUNG: Zielattribut als absolutes Wert.	57
TABELLE: Verteilung von Beispielen in 9 Klassen.	58
ABBILDUNG: Zielattribut als Matrix 3X3.	58
ABBILDUNG: Verteilung in 3 Zielklassen.	59
ABBILDUNG: Zielattribut als Funktion von Auftragswert und dem Gewinnin %.	60
ABBILDUNG: Verteilung in 6 Zielklassen.	61
ABBILDUNG: Zielattribut als Funktion von Auftragswert und dem Gewinnin %.	62
TABELLE: Vorhersage von 10% schlechtesten Baustellen.	63
ABBILDUNG: Definition von Top&Flop Baustellen.	64
ABBILDUNG: Vorteile von SVM.	66
ABBILDUNG: Verfahren im Kepler.	70
ABBILDUNG: Visualisierung von Daten im Kepler.	71
ABBILDUNG: Visualisierung von Ergebnissen.	72