

Fallstudien/Themen: CPM Sommer 2018

Basis für den praktischen Teil der nachfolgenden Referatsthemen ist der MS SQL-Server in Verbindung mit Visual Studio (Data Tools) und Excel. Vorgeschlagenen Datenquellen: MS Beispiel-Datenbanken AdventureWorks(DW), Libri Daten (DMC 2010) oder Absatzprognose Online-Shop (DMC 2012). Es steht Ihnen frei, neben dem SQL Server weitere SW-Produkte als Data Mining Tools einzusetzen. Dabei ist jeweils eine prototypische Lösung der Aufgabenstellung mit Hilfe dieser Werkzeuge und der vorgegeben Datenbasis zu erstellen. Gelegentlich kann es vorkommen, dass die vorgegeben Daten für bestimmte Analysen ungeeignet sind; in diesem Fall sind die Daten geeignet zu ergänzen oder durch andere Datenquellen zu ersetzen und die aufgetretene Problematik herauszuarbeiten. Jede Fallstudie umfasst betriebswirtschaftliche und technologische Aspekte.

Wir erwarten von Ihnen auch Eigeninitiative und eine methodisch-analytische Bearbeitung des Themas.

Themen:

1. Vom operativen Datenbestand zum Data Warehouse (DW)

Untersuchen Sie auch theoretisch: den Extraktionsprozess mit Filterung, Bereinigung, Transformation, Harmonisierung und Verdichtung der Daten.

Anleitung: Wählen Sie relevante Daten aus der AdventureWorks Datenbank aus und bereinigen Sie diese. Setzen Sie die Verkaufszahlen mit volkswirtschaftlichen Kenngrößen in Bezug. Die dafür geeigneten Zahlen sollten Sie aus öffentlichen Datenquellen im gleichen Zeitraum über das Internet beziehen und mit dem Transformation Service importieren. Achten Sie auf gleiche Semantik der Daten (Harmonisierung). Modellieren Sie Ihre Daten in einem Star-Schema. Berechnen Sie Aggregatwerte und fügen Sie diese Ihrem Data Warehouse hinzu.

Datenquelle: AdventureWorks (MS SQL Server 2014), öffentliche Internetquellen

2. Alternative Modellierung von DW-Lösungen

Schlagen Sie ausgehend von der AdventureWorks Datenbank und der Aufgabe 1 weitere mögliche Strukturierungen vor und realisieren Sie diese mit dem SQL-Server. Untersuchen Sie auch theoretisch Star-, Snowflake- und Galaxie-Modell.

Anleitung: Modellieren Sie die Daten aus Aufgabe 1 (bzw. Ihre eigene Datenauswahl) alternativ als Snowflake- und als Galaxie-Modell. Dazu sind die Dimensionen zu normalisieren und die Fakten aufzuspalten. Diskutieren Sie die Modelle und ihre Konsequenzen praktisch und theoretisch

Datenquelle: AdventureWorks (MS SQL Server 2014)

3. Vom DW zum OLAP-Modell

Entwickeln Sie mit Hilfe des Cube Editors im Business Intelligence Development Studios (BID Studio) eine mehrdimensionale Datenstruktur zur Untersuchung der Bestellungen hinsichtlich Umsatz und Anzahl Artikel in den Dimensionen Zeit, Artikel, Kunden und Verkäufer. Um welchen Faktor vergrößert sich der Datenbestand durch die Generierung der Aggregatwerte? Welchen Einfluss haben verschiedene Speichermodelle?

Untersuchen Sie auch theoretisch verschiedene Speichermodelle (MOLAP, ROLAP, HOLAP) und das Dimensional Fact Modeling.

Datenquelle: AdventureWorks (MS SQL Server 2014)

Anleitung: Wählen Sie die notwendigen Daten aus dem DW aus. Erstellen Sie mit Hilfe des BID Studios einen Datenkubus, mit dem die Fragestellung sich beantworten lässt. Wählen Sie geeignete Attribute für die Aggregatbildung der Dimensionen.

4. Objektorientiertes Modell als Alternative zum OLAP-Modell

Entwerfen Sie ein objektorientiertes multidimensionales Datenmodell und realisieren Sie die notwendigen Basisklassen in db4o.

Anleitung: Entwerfen Sie zunächst die fachlichen Klassen für die Dimensionen und die Strukturklassen Zelle und Kubus als Behälter der Fakten. Implementieren Sie die Klassen in Java (z.B. mit Eclipse). Dann übernehmen

WIM 2: Corporate Performance Management (Roth, Laux)

Sie diese in die oo-Datenbank db4o (open source). Schreiben Sie ein kleines Programm, das die Daten (über ODBC) aus der Datenbank des SQL-Servers liest und als Objekte in db4o instanziiert.

5. Zeitreihenanalyse:

Prognostizieren Sie aus der Umsatzhistorie (Tag 1 – Tag 42) die Bestellungen von Tag 43 - 56.

Anleitung: Importieren Sie die Daten in das Business Intelligence Developer Studio zur Zeitreihenanalyse.

Erstellen Sie die Prognose mit Hilfe der dort verfügbaren Verfahren (ARIMA, ARTXP). Untersuchen Sie auch gleitender Mittelwerte, Exponentieller Glättung und einen linearen Trend mit saisonalem Einfluss. Sie können hierzu auch Excel mit seinen Statistik Funktionen verwenden. Vergleichen Sie Ergebnisse diskutieren Sie die verschiedenen Verfahren. Überprüfen Sie die Ergebnisse mit alternativen Zeitreihentools.

Datenquelle: Absatzprognose Online Shop

6. Data Mining (Clustering):

Wie können die Produkte aus den vorgegebenen Online Shop Daten sinnvoll gruppiert werden?

Anleitung: Zur automatischen Gruppierung von Kunden werden gerne Clustering-Verfahren eingesetzt.

Beschreiben Sie theoretisch das Konzept und erklären Sie die Begriffe des Clustering. Überlegen Sie sich Kriterien für die Ähnlichkeit (Distanzfunktion) und untersuchen Sie mindestens 2 Verfahren zur Einteilung der Produkte theoretisch. Wenden Sie dann beide Algorithmen auf die Produkte an und vergleichen Sie die Ergebnisse.

Datenquelle: Absatzprognose Online Shop

7. Decision Support System (DSS):

Wie entwickelt sich der Umsatz und das Ergebnis pro Region/pro Verkaufsbeauftragter, wenn sich die Preise um x% verändern? Zeigen Sie anhand eines Beispiels auf, wie sich solche Szenarien grundsätzlich modellieren und realisieren lassen? Untersuchen Sie auch theoretisch welche Implementierungskonzepte für schwach belegte Speicher (Sparse Data) existieren. Welche Konzepte realisieren die MS Analysis Services?

Hilfsmittel: SQL-Server 2014 und ggfs. MS-Excel

Datenquelle: AdventureWorks (MS SQL Server 2014)

8. EIS/MIS: Berichterstattung:

Zeigen Sie anhand einer prototypischen Präsentation auf, wie eine Standardberichterstattung und eine Ad-Hoc-Berichterstattung aussehen können? Welche Funktionalitäten sind bei unterschiedlichen Berichtstypen jeweils erforderlich?

Die Hilfsmittel sind frei wählbar. Vorschlag: Arcplan, Crystal Reports oder Reporting Services

Datenquelle: AdventureWorks (MS SQL Server 2014)

9. Plan-/Ist-Vergleiche (Abweichungsanalyse)

Entwerfen Sie ein Konzept und eine praktische Lösung wie Plan-/Ist-Abweichungen von Daten auf verschiedenen Aggregationsstufen identifiziert, analysiert und aggregiert berichtet werden können.

Berücksichtigen Sie hierbei auch dass sich die Gliederungsstrukturen in Unternehmen (Unternehmensorganisation, Produktstruktur, Vertriebsstruktur etc.) im Zeitablauf verändern können.

Die Hilfsmittel sind frei wählbar.

Datenquelle: AdventureWorks (MS SQL Server 2014)

10. Datenvorbereitung zum Data Mining

Die Datenvorbereitung umfasst die Kontrolle der Datenqualität einschließlich der Korrektur widersprüchlicher und fehlerhafter Daten. Fehlerhafte Daten können syntaktisch durch Datentypverletzung und semantisch durch inhaltliche Widersprüche (z.B. Audio Download und Gewicht > 0) erkannt werden. Abhängig vom Data Mining Verfahren kann eine Datenanalyse nur dann durchgeführt werden, wenn die Datencodierung dem Verfahren angepasst wird. Manche Algorithmen können z.B. nur mit numerischen Werten umgehen, andere erfordern diskrete oder binäre Werte. Daher kann es notwendig sein, dass kontinuierliche Werte diskreditiert werden oder diskrete Wert müssen in mehrere binäre Attribute zerlegt werden. Z.B. das Einkommen muss einer

WIM 2: Corporate Performance Management (Roth, Laux)

Einkommensgruppe (Intervall) zugeordnet werden. Erstellen Sie exemplarisch für verschiedene Szenarien Empfehlungen, wie

- widersprüchliche Daten zu behandeln sind
- fehlende Daten ergänzt werden können
- kontinuierliche Zahlen diskreditiert werden können
- diskrete Werte in binäre Attribute zerlegt werden können

Hilfsmittel: frei wählbar (ggfs. Analysis Server, Excel)

11. Hauptkomponentenanalyse als Vorbereitung zum Data Mining

Bestimmte Algorithmen (z.B. Bayes Klassifikation) gehen von unabhängigen Einflussgrößen aus. Daher müssen abhängige Attribute so transformiert werden, dass die neuen Attribute unabhängig voneinander werden. Das Verfahren der Wahl ist hierfür die Hauptkomponentenanalyse. Untersuchen Sie mit Hilfe dieser Analyse die Libri Daten hinsichtlich ihrer Abhängigkeiten untereinander. Zeigen Sie exemplarisch die Wertereduktion durch die Hauptkomponentenanalyse. Wie können abhängigen Attributen in unabhängige Komponenten umgewandelt werden? Welche Schlüsse ziehen sie aus den ermittelten Hauptkomponenten?

Hilfsmittel: frei wählbar (ggfs. Analysis Server, Excel)

12a. Data Mining (Kategorisierung, Entscheidungsbaum):

Entwickeln Sie eine automatische Klassifikation der Artikel für die Absatzmengen.

Anleitung: Entwickeln Sie einen Klassifikationsbaum und stellen Sie die zugehörigen Klassifikationsregeln auf, die auf einer Trainingsmenge von 10% der Artikel beruhen. Bewerten Sie die Klassifikation auf Grund ihres Klassifikationsfehlers, wenn Sie hierzu weitere 10% Testdaten verwenden. Vergleichen Sie die Ergebnisse mit den realen Werten der Prognose (dmc2012_real.txt)

Datenquelle: Absatzprognose Online Shop

12b. Data Mining (Kategorisierung, Naïve Bayes Klassifikator):

Entwickeln Sie eine automatische Klassifikation der Artikel für Absatzmengen.

Anleitung: Verwenden Sie den „naiven Bayes-Algorithmus“ um den Entscheidungsbaum aufzubauen. Die statistische Basis soll auf einer Trainingsmenge von 10% der Artikel beruhen. Bewerten Sie die Klassifikation auf Grund ihres Klassifikationsfehlers, wenn Sie hierzu weitere 10% Testdaten verwenden. Vergleichen Sie die Ergebnisse mit den realen Werten der Prognose (dmc2012_real.txt)

Datenquelle: Absatzprognose Online Shop

13. Data Mining (Neuronales Netzwerk):

Prognostizieren Sie aus der Umsatzhistorie (Tag 1 – Tag 42) die Bestellungen von Tag 43 - 56.

Anleitung: Verwenden Sie die vorhandenen Verfahren des Business Intelligence Developer Studios oder Szenario von Cognos zur Definition des Künstlichen Neuronales Netzes (KNN). Trainieren Sie dieses mit den Bestelldaten im Zeitraum Tag 1 – Tag 42. Dann wenden Sie das Netzwerk auf den Prognosezeitraum an. Vergleichen Sie die Ergebnisse mit denen von Aufgabe 5. Beschreiben Sie auch die Elemente und Algorithmen der KNN theoretisch.

Datenquelle: Absatzprognose Online Shop

14. Vorhersage numerischer Werte (Regressionsbaum):

Erstellen Sie einen Regressionsbaum zur Verkaufsprognose. Als Trainingsdaten stehen die Absatzzahlen der Produkte im Zeitraum Tag 1 – Tag 42 zur Verfügung. Der Baum sollte bezüglich der Hierarchiestufen minimiert werden und die Vorhersagefunktion sollte linear sein. Wenden Sie Ihre Lösung auf folgende Fragestellungen an: Welche Subkategorien erzielen die größten Stückzahlen? Wie werden sich die zukünftigen Verkäufe (Tag 43 – Tag 56) entwickeln.

Datenquelle: AdventureWorksDW (MS SQL Server 2014)

Anleitung: Die Ausarbeitung soll eine Darstellung der Theorie der Entscheidungsbäume für numerische Vorhersagen enthalten. Illustrieren Sie an der gestellten Aufgabe die Wahl der optimalen Klassifizierer (hier

Produktpreis). Stellen Sie dar, wo und wie das Abschneiden (pruning) von Zweigen der Hierarchie erfolgt und wie Sie die Vorhersagemodelle (lineare Funktionen) ermitteln.

15. Approximative Lösung durch genetische Algorithmen:

Erstellen Sie ein Exzerpt der Ideen und Konzepte von genetischen Algorithmen. Entwickeln Sie ein einfaches Modell (Vererbung, Mutation) mit einer Fitnessfunktion für die Lösung des „Travelling Salesman“ Problems

Datenquelle: frei wähl- und gestaltbar, z.b. Geodaten von Google Maps oder OSM

Anleitung: Die Ausarbeitung soll eine Darstellung der Prinzipien und Konzepte von genetischen Algorithmen enthalten. Ein möglicher Algorithmus soll dargestellt und exemplarisch realisiert werden. Die Wirksamkeit und Effizienz sollen an Hand der gestellten Aufgabe theoretisch und praktisch bewertet werden.

16. Ermittlung und Präsentation von 'Softfacts'

Entwerfen Sie ein Konzept und eine praktische Lösung, wie so genannte weiche Faktoren z.B.

Mitarbeiterzufriedenheit, Kundenzufriedenheit mittels Webformular erfasst, in das Data-Warehouse gespeichert und mittels OLAP ausgewertet werden können. Problemstellung u.a. : Umgang mit Texteinträgen, Aggregation von erhobenen Daten, Bewertung freier Einträge, Ordnungsskala. Die Hilfsmittel sind frei wählbar.

Datenquelle: Adventure Works

17. Balanced Score Card

Entwerfen Sie ein Konzept und eine prototypische Lösung für eine Unternehmenssteuerung mittels einer Balanced Score Card. Erklären Sie das theoretische Konzept. Entwerfen und modellieren Sie die erforderlichen Daten. Erstellen Sie einen Prototyp mit ausgewählten Kennzahlen und zeigen, wie ein Unternehmen damit beurteilt und gesteuert werden kann.

Datenquelle: frei wählbar; u.a. AdventureWorksDW (MS SQL Server 2014)

18. SAP Business Information Warehouse

Zeigen Sie auf, wie das Standardprodukt SAP BW den Prozess von der operativen Datenhaltung bis zur Managementinformation löst. Geben Sie einen Überblick über das Gesamtkonzept und verdeutlichen Sie dieses anhand eines beispielhaften Prototyps. Zeigen Sie die Stärken und Schwächen dieses Konzeptes auf.

Datenquelle: frei wählbar

=> Aufsplittung des Themas in 3 Teilthemen -> 3 Bearbeiter:

a) ETL-Prozess und Datenspeicherung
Ausarbeitung, Präsentation, Beispiel

b) Query, Reporting und Analyse
Ausarbeitung, Präsentation, Beispiel

c) Business Analytics und Data Mining
Ausarbeitung, Präsentation, Beispiel

19. Customer Relationship Management

Bitte bearbeiten Sie bei Ihrer Ausarbeitung insbesondere folgende Themenfelder:

- Theoretisches Grundkonzept und Zielsetzung des CRM
 - Konzeption und Modellierung eines CRM-Systems
 - Welche Funktionalitäten und Berichte halten Sie für erforderlich?
 - Prototypische Realisierung eines CRM-Systems basierend auf der
- Datenquelle: AdventureWorksDW (MS SQL Server 2014)
- Zeigen Sie Stärken und Schwächen dieses Konzeptes auf

20. Warenkorbanalyse

Untersuchen Sie das Käuferverhalten hinsichtlich der Waren, die bei einem Einkauf gemeinsam erworben werden. Korrigieren Sie für Ihre Analyse ggf. die SalesFact-Daten so, dass TimeID, TerritoryID und CustomerID eindeutig einen Einkauf identifizieren. Untersuchen Sie auch den Einfluss von "A-Priori"-Algorithmen sowohl theoretisch als auch durch eigene Implementierung und stellen Sie die Ergebnisse den SQL-Abfragen auf dem SQL-Server gegenüber.

Datenquelle: AdventureWorks (evtl. Daten ergänzen bzw. korrigieren)

21. Corporate Performance Management (CPM)

Was versteht man unter CPM? Stellen Sie das theoretische Konzept vor? Was ist an diesem Konzept neu im Vergleich zu den bisherigen Business Intelligence-Konzepten? Stellen Sie anhand einer prototypischen Präsentation vor, wie dieses Konzept in die Praxis überführt werden kann. Welches Fazit (Vor- und Nachteile) können Sie ableiten?

Datenquelle: AdventureWorks (evtl. mit Daten zu ergänzen)

Hilfsmittel: frei wählbar (ggfs. Analysis Server, Excel)

22. Clustering und Portfoliotechnik - Methoden des Data Mining und der Dateninterpretation

Stellen Sie die Verfahren Clustering und Portfoliotechnik mit ihren jeweiligen Stärken und Schwächen vor. Inwiefern kann das Clustering verwendet werden um die Portfoliobildung zu unterstützen? Sind die beiden Verfahren kombinierbar? Welchen Mehrwert und welche Probleme ergeben sich durch die Kombination der beiden Verfahren? Unterlegen Sie Ihre Überlegungen anhand der Durchführung der Verfahren in einem exemplarischen Beispiel.

Datenquelle: frei wählbar, Hilfsmittel: frei wählbar

23. Business Case für Führungsinformationssysteme

Wie kann ein Business Case für FIS-Projekte aufgebaut werden? Welche Kriterien gilt es zu berücksichtigen? Wie kann der Nutzen dargestellt und ex-ante bewertet werden? Entwerfen Sie ein Konzept und Implementieren Sie eine prototypische Lösung zur Business Case-Erstellung mit Hilfe multidimensionaler Analysen (OLAP). Demonstrieren Sie dieses anhand eines exemplarischen Beispiels.

24. Währungsumrechnung

Wie wird dieses Thema in Bezug auf schwankende Wechselkurse und unterschiedliche Berichtswährungen betriebswirtschaftlich behandelt? Wie kann man diese Problematik mit Hilfe des SQL-Servers behandeln und für eine OLAP-Auswertung aufbereiten.

25. Business Process Management (BPM)

Was versteht man unter BPM? Stellen Sie das theoretische Konzept vor? Wie lässt sich dieses Konzept von Corporate Performance Management (CPM) abgrenzen? Welche Verbindung zu CPM lässt sich herstellen? Was ist an diesem Konzept neu im Vergleich zur bisherigen Geschäftsprozessoptimierung? Inwiefern eignet sich dieses Konzept zur Unternehmenssteuerung? Stellen Sie anhand einer prototypischen Präsentation vor, wie dieses Konzept in die Praxis überführt werden kann. Welches Fazit (Vor- und Nachteile) können Sie ableiten?

Datenquelle: AdventureWorks (evtl. mit Daten zu ergänzen)

Hilfsmittel: frei wählbar (ggfs. Analysis Server, Excel)

26. BI für Social Media Daten

Wie können Social Media Daten in BI-Systemen ausgewertet werden? Beschreiben sie verschiedene Techniken zur Datensammlung. Welche Anforderungen sind an eine Klassifizierung von Einträgen zu stellen. Wie könnte ein Monitoring für die unstrukturierten Daten aussehen?

27. Maschinelles Lernen

Auf welchem Prinzip basiert maschinelles Lernen. Stellen sie die gängigen Konzepte des maschinellen Lernens vor. Arbeiten Sie die Unterschiede zwischen überwachtem und unbeaufsichtigtem Lernen heraus. Wie ordnen sich konkrete Verfahren aus der Vorlesung ein? Was ist eine Ground truth? Wie können Ergebnisse aussehen? In welchen Situationen lohnt es sich besonders? Wo sind die Grenzen und wo versagen maschinelle Lernverfahren?

Untersuchen Sie, wie maschinelles Lernen in einem Betrieb sinnvoll eingesetzt werden kann und stellen Sie Ihr Konzept sowie einen Prototypen vor.

28. MapReduce Programmiermodell

Was versteht man unter MapReduce? Welche Aufgaben(typen) eignen sich für dieses Konzept? Was versteht man unter Map, Shuffle, Reduce und Combine Phasen?

Implementieren Sie einen Business Case mit Hadoop, welcher die Vor- und Nachteile dieses Konzeptes illustriert.