Part X

Business Intelligence Anwendungen

Business Intelligence Anwendungen

- Definition
- Use Cases
- Report & BSC

Definition

Business Intelligence

Diverse definitions:

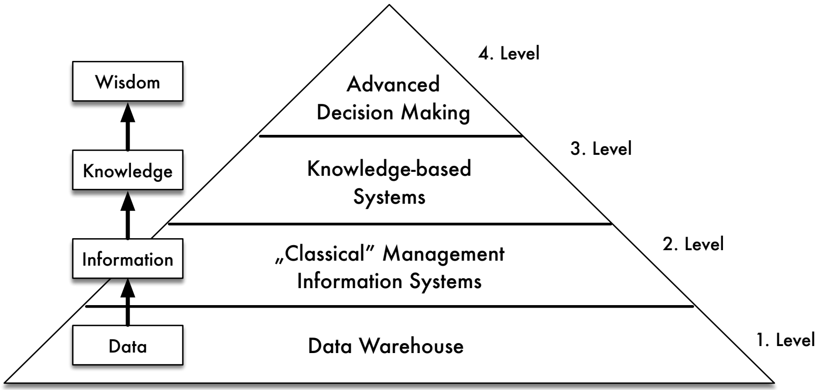
- 1989 term Business Intelligence coined [Dresner 1989]
- from the 60's (since data processing):
  - Management Information Systems
  - Management Support Systems
  - Executive Information Systems
- Differentiation:
  - In a narrower sense
  - Analysis-oriented
  - In a broader sense

Intelligence

Terminology:

- Finding orders,
- Rules for commonalities (consilience),
- Rules for co-occurence and sequential occurrences of events,
- Targeted collection and transfer of information,
- Information logic

Knowledge Pyramid



Business Intelligence

- Data- and information processing for the management
- Information logistics: filtering of information
- MIS: fast and flexible evaluations
- Early warnings in companies ("Alerting")
- BI = Data Warehousing
- Information and knowledge storage
- Prozess of gathering $\rightarrow$ Diagnosis $\rightarrow$ Therapy $\rightarrow$ Forecast $\rightarrow$ Control

[Mertens 2002]

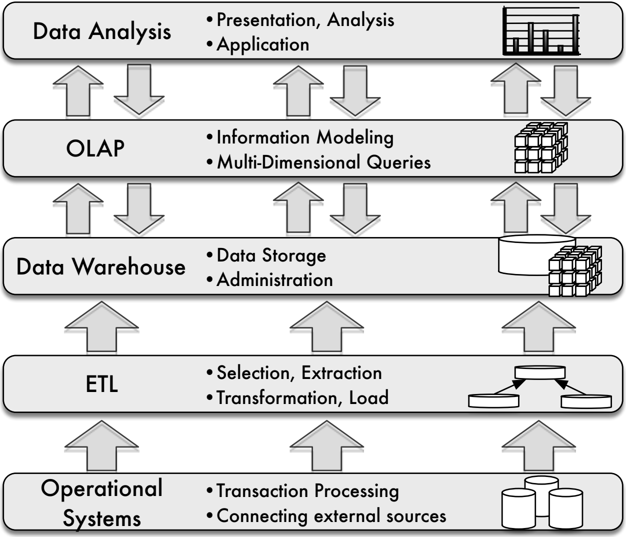
Business Intelligence

Business Intelligence is the analytical prozess, that transforms – fragmented – company and competition data in action-oriented knowledge about skills, positions, actions and targets in the regarded internal or external fiels of action (actors and processes).

[Grothe & Gensch 2000]

- Analytical process: planning, deciding and directing
- Omniscient data integration and provision
- Action-oriented knowledge: communication + information + knowledge representation

Business Intelligence Prozess



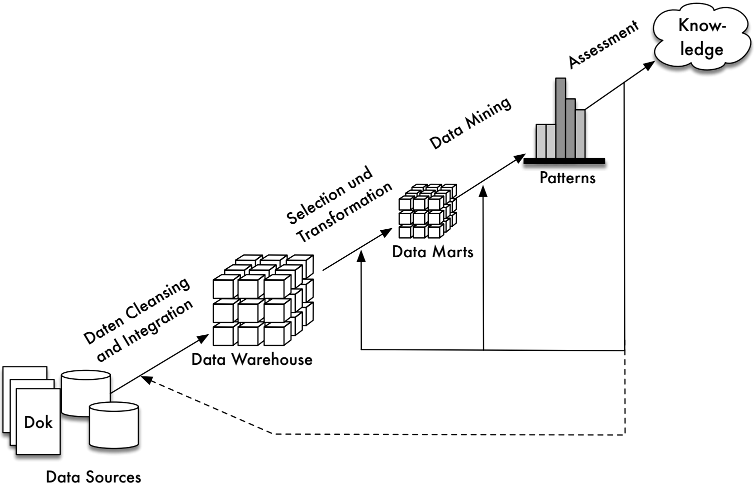
Data Warehouse and Business Intelligence

- Data Warehouse is a central information storage
- BI: methods to connect quantitative, qualitative, internal and external information
- DW data needs to be accordingly filtered and aggregated to represent personalized information / knowledge
- Data Mart is starting point for domain-specific analysis

Large data volume:

- Data in the OLAP area grows permanently
  $\hookrightarrow$ Overview of structure in the data by exploratory methods
- Data Mining pattern recognition

Knowledge Discovery Prozess



[Han & Kamber 2006]

Business Intelligence

Business Intelligence is the decision-oriented collection, preparation and presentation of business-relevant information.

[Schrödl 2006]

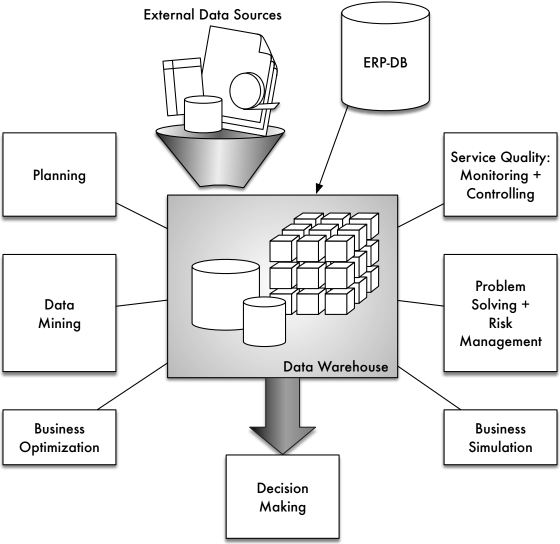
- Improve decision basis,
- Data collection: heterogeneous sources and requirements (e.g. security)
- Transform raw data in information (e.g, mathematical, rule based)
- Information representation for the user
- Concentrate on business relevance (optimization benefits & efforts)

BI Cycle

1. Quantification and qualification of business information
2. Analysis of the obtained data
3. Gaining insights supporting business processes
4. Evaluating the insights given the goals
5. Implementing the relevant insights in concrete actions

[Vitt et al. 2002]

Business Intelligence



Use Cases

Typical DW Use Cases

- Which clients do we have?
  ↪ Customer Relationship Management
- How do our costs develop?
  ↪ Supply Chain Management
- Where is further potential in our product range?
  ↪ Customer Basket Analysis
- ...

Typical Data Mining Methods

- Association rules – What has been bought together in a customer basket?
- Classification approaches – Which customer groups shall get special offers?
- Clustering – Which commonalities exist between our clients / suppliers?
- ...

Customer Basket Analysis

- Transactions at a coubter (transaction data base):
  - T1: {Müller-Thurgau, Riesling, Dornfelder}
  - T2: {Riesling, Erfurter Bock, Ilmenauer Pils, Anhaltinisch Flüssig}
  - T3: {Müller-Thurgau, Riesling, Erfurter Bock }
- Customer basket analysis: Which products are bought frequently together?
- Targets:
  - Optimization Shop Layout
  - Cross-Marketing
  - Add-On Sales

Association Rules

- Rule type:
  Body $\rightarrow$ Head [support, confidence]
- Example:
  - buys(X, "Red wine") $\rightarrow$ buys(X, "Erfurter Bock") [0.5%, 60%]
  - 98% of all clients buying Müller-Thurgau and Riesling pay by credit card.

Basic Definitions

according to [Agrawal und Srikant (1994)]

- Items $I = \{i_1, i_2, \ldots, i_m\}$ – Population of literals
- Itemset X: $X \subseteq I$
- Database D – Set of transactions $X \subseteq I$
- $X \subseteq T$
- Lexikographical sorting in T and X
- Length k of a itemset: number of elements
- k-Itemset: Itemset of length k

Basic Definitions (2)

- Support of the set X in D: share of transactions in D, that contain X:
  $supp(X) = \frac{|X|}{|D|}$
- Association rule: $A \rightarrow B$, with $A \subseteq I$, $B \subseteq I$ and $A \cap B = \emptyset$
- Support s of a association rule $A \rightarrow B$ in D: $s = supp(X \cup Y)$
- Confidence c of a association rule $A \rightarrow B$ in D: share of transactions, that contain B when they are present in A –
  $c = conf(B|A) = \frac{supp(A \cup B)}{supp(A)}$

Problem: Identify all association rules that in D exhibit a support $\geq$ minsup and a confidence $\geq$ minconf.

Example Association Rules

minsup = 20 %

TID	Items
1	Erfurter Bock, MT, Riesling
2	Erfurter Bock, MT, Dornfelder
3	Ilmenauer Pils, MT
4	Anhaltinisch Flüssig, Dornfelder, Riesling
5	Berliner Bräu, Dornfelder, Riesling
6	Kölnische Weisse, MT
7	Anhaltinisch Flüssig, Dornfelder

- $supp(\textit{MT}) \approx 57\%$
- $supp(\textit{Riesling}) = supp(\textit{Dornfelder}) \approx 43\%$
- $supp(\textit{Erfurter Bock}) = supp(\textit{Anhaltinisch Flüssig}) \approx 29\%$
- $supp(\textit{Ilmenauer Pils}) = supp(\textit{Berliner Bräu}) = supp(\textit{Köln. Weisse}) \approx 14\%$.
- potential candidates: MT, Riesling, Dornfelder, Erfurter Bock, Anhaltinisch Flüssig

Example Association Rules (2)

- possible combinations of all candidates:

Itemset	Support in %
(Erfurter Bock, MT)	$\approx$ 29
(Erfurter Bock, Riesling)	$\approx$ 14
(Erfurter Bock, Dornfelder)	$\approx$ 14
(Erfurter Bock, Anhaltinisch Flüssig)	0
(MT, Riesling)	$\approx$ 14
(MT, Dornfelder)	$\approx$ 14
(MT, Anhaltinisch Flüssig)	0
(Riesling, Dornfelder)	$\approx$ 29
(Riesling, Anhaltinisch Flüssig)	0
(Dornfelder, Anhaltinisch Flüssig)	$\approx$ 29

Apriori Algorithm

Input I, D, minsup
Output $\bigcup_k L_k$
C_k: candidates that shall be counted of length k
L_k: set of all frequent occurring itemsets
of length k
initialize L_1:= 1-itemsets of I, k:= 2
WHILE $L_{k-1} \neq \emptyset$
C_k := AprioriCandidateGeneration(L_{k-1});
FOR EACH Transaction $T \in D$
CT := Subset(C_k, T)
// all candidates from C_k, that T contains
FOR each candidate $c \in CT$ c.count++
L_k := $\{c \in C_k | (c.count/|D|) \geq minsup\}$
k++

Improving the Efficiency of the Apriori Algorithm

- Counting the support using a hash table
  - [Park, Chen, Yu 1995]
  - Hash tabele instead of a hash tree
  - k-itemset, whose bucket has a numerator smaller than the minimal support, cannot be frequent
    more efficient access to candidates, less accurate computation
- Transaction Reduction
  - [Agrawal & Srikant 1994]
  - Transactionens, that do not have a k-frequent itemset are redundant, i.e., they can be removed
  - Database scan is more efficient, but there is writing effort

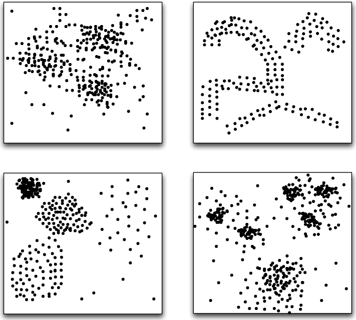
Improving the Efficiency of the Apriori Algorithm (2)

- Partitioning
  - [Savasere, Omiecinski & Navathe 1995]
  - Itemset only frequent when it is frequent in a partition
  - Exploiting the main memory (Partition)
  - Partition efficient, but effort for merging
- Sampling
  - [Toivonen 1996]
  - Application of Apriori on an excerpt (Sample)
  - Counting of found rules on the whole database

Cluster Approaches

- Identification of a finite set of groups in the data → Search for partitioning
- Similarity within a group
- Preferably significant difference between the groups

Occurring patterns (size, form, density):



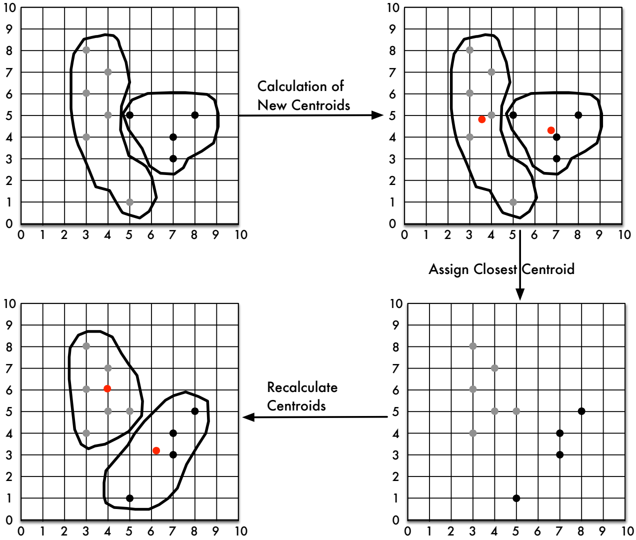
Distance functions

- Similarity metric $sim(objekt_1, objekt_2)$
- Distance function $dist(objekt_1, objekt_2)$ $O \times O \rightarrow R_+$
  - small distance $\rightarrow$ similar, large distance $\rightarrow$ not similar
  - $dist(objekt_1, objekt_2) = 0$, given if $objekt_1 = objekt_2$
  - Symmetry: $dist(objekt_1, objekt_2) = dist(objekt_2, objekt_1)$
  - For metrics:
    $dist(objekt_1, objekt_3) \leq dist(objekt_1, objekt_2) + dist(objekt_2, objekt_3)$

Partitioning Clustering

Clustering through minimizing variance
Input: Tuple set D, numer of classes k
Output: Cluster C
Create an initial partitioning of D in k classes
Compute set $C^* = \{C_1, \ldots, C_k\}$ of
    centroids per class
$C := \{\}$
repeat
    C := C^*
    Partition: Create k classes by assigning
    each point to the closest centroid from C
    Compute centroids: Calculate the set
    $C^* = \{C_1^*, \ldots, C_k^*\}$ of centroids
    for the newly determined classes
until C = C^*

Cluster Approaches: Illustration



Advantages and Disadvantages

Advantages:

- linear effort per iteration, few iterations
- easy to implement
- k-means [MacQueen 1967]: most popular clustering algorithm

Disadvantages:

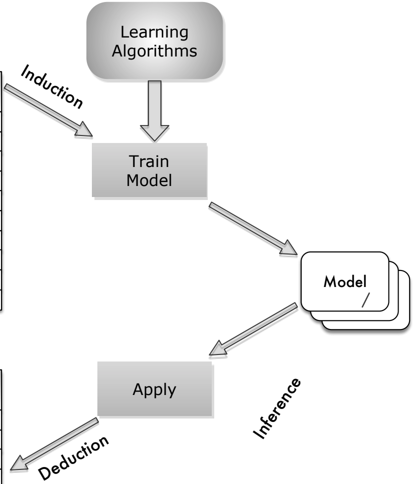
- Sensitive to noise and outliers
- Convex form of the clusters
- Fixed number of clusters
- Initial distribution important for runtime and end result

Classification: Example

Do we like the Wine?

TID	Wine Color	Res. Sugar g/l	Alcohol	Class
1	White	18	10	Yes
2	Red	20	9	Yes
3	Rose	22	9	No
4	Rose	15	8	No
5	Red	30	5	Yes
6	White	18	10	Yes
7	Red	15	15	No
8	White	45	5	Yes
9	White	18	14	Yes
10	Red	8	10	No

Training Set



TID	Wine Color	Res. Sugar g/l	Alcohol	Class
11	Red	23	10	?
12	Rose	15	12	?
13	White	22	10	?
14	White	30	6	?
15	Red	12	14	?

17	White	5	16	?

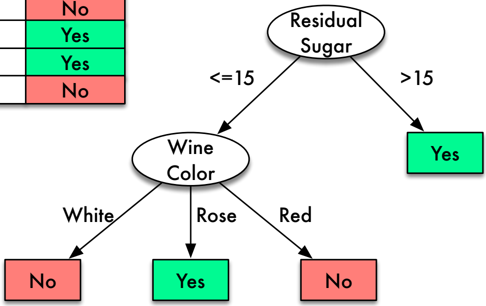
Test Set

Classification

- Given is a set of object with attributes $o = (x_1, \ldots, x_d)$ and their membership to the set of classes C
- Search for classifier K for new objects $\rightarrow K : Objekts_{new} \rightarrow C$
- Class membership a-priori known $\rightarrow$ Difference to clustering approaches
- Similar to forecast (e.g., linear regression)

Classification Result

TID	Wine Color	Res. Sugar g/l	Alcohol	Yes/No
1	Red	23	12	Yes
2	White	15	10	No
3	Rose	14	10	Yes
4	White	30	6	Yes
5	Red	12	14	No



Classification Quality

		Forecast	
		Member of class	Not member of class
True labels	Member of class	True Positive	False Negative
	Not member of class	False Positive	True Negative

- Accuracy: $\frac{TP+TN}{TP+FN+FP+TN}$
- Precision: $p = \frac{TP}{TP+FP}$
- Recall: $r = \frac{TP}{TP+FN}$
- F-Measure: $F = \frac{2 \cdot TP}{2 \cdot TP+FN+FP}$

Classification Methods

- Decision Tree
- Rule-based
- Linear discriminant analysis by Fisher
- Categorical regression, Log-Linear models
- Neural networks
- Naive Bayes and Bayesian Belief Networks
- Support Vector Machines

Decision Tree

- Process: Splitting and Partitioning
- Explicites knowledge is found
- Easy to understand
- Easy to visualize

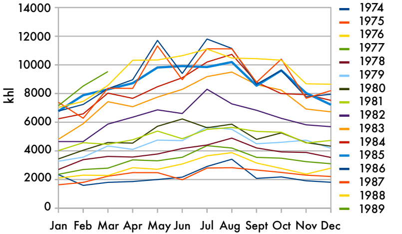
Algorithm for the Decision Tree

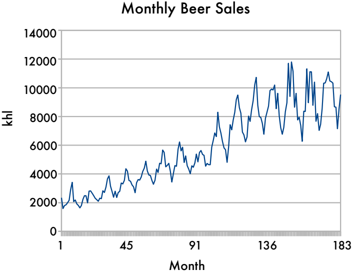
```
Input:  Training data
Initialization:  all data points (instances)
   belong to the root node
WHILE Split attribute exists OR data points
of a node in different classes
   Choose a split attribute (Splitting Strategy)
   Partition data points of a node
   according to the attribute
   Recursion for all partitions
```

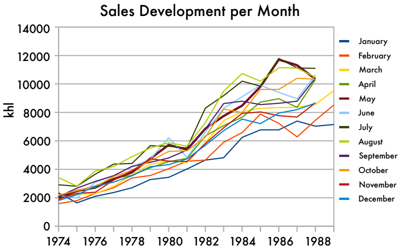
Forecast: Example

Monthly Beer Sales of a Brewery (khl)

1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
2339	1638	2101	2363	2697	3279	3438	4021	4646	4811	6236	6770	6771	7386	7034	7150
1588	1798	2307	2700	3388	3561	4044	4570	4646	5896	6582	7881	7237	6279	7449	8525
1800	2235	2281	2794	3609	4343	4584	4461	5868	7426	8029	8290	8335	8370	8569	9530
1858	2481	2827	3371	3570	4103	4536	4771	6346	7076	7661	8720	8966	8356	10320	
2001	2479	2713	3303	3783	4749	5711	5383	6857	7749	8471	9813	11709	11318	10340	
2169	1988	3083	3555	4163	4711	6225	4843	6602	8293	9103	9913	9402	8964	10641	
2911	2804	3657	4364	4405	5661	5609	5504	8295	9183	10198	9847	11799	11119	11100	
3414	2820	3872	4198	4890	5503	5860	5633	7278	9496	10725	10196	11147	11113	10474	
2077	2666	3149	3547	4206	4494	4800	5360	6829	8620	8785	8546	8645	8783	10427	
2184	2494	2773	3491	3923	4595	5256	5297	6269	8237	7994	9613	9615	10397	10329	
1913	2308	2382	3246	3893	4740	4576	4546	5814	6919	7929	8038	7765	7672	8677	
1809	2212	2798	3102	3543	4179	4330	4733	5686	6721	7527	7217	7948	8202	8651	

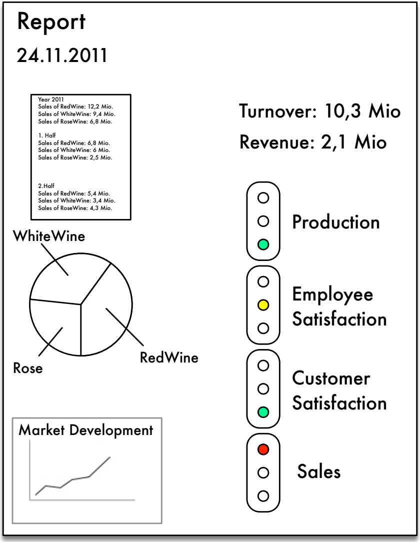




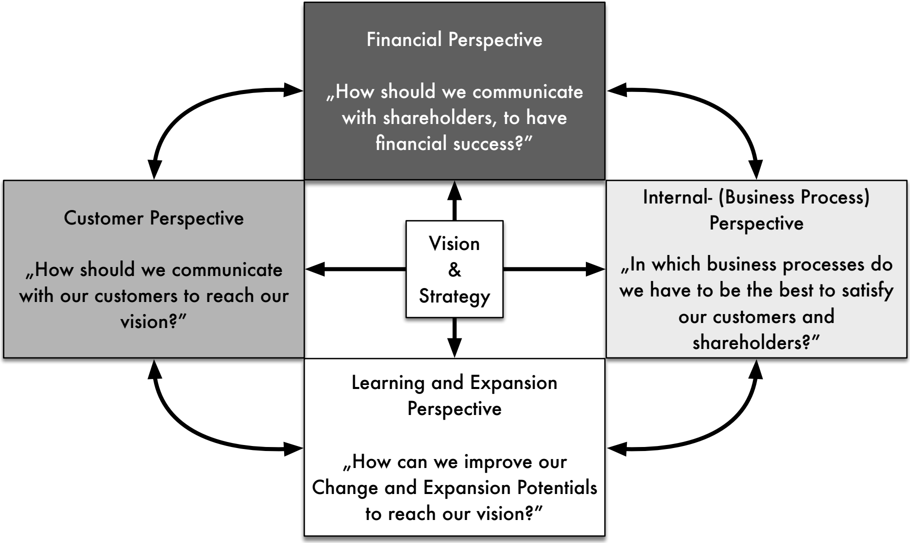


Report & BSC

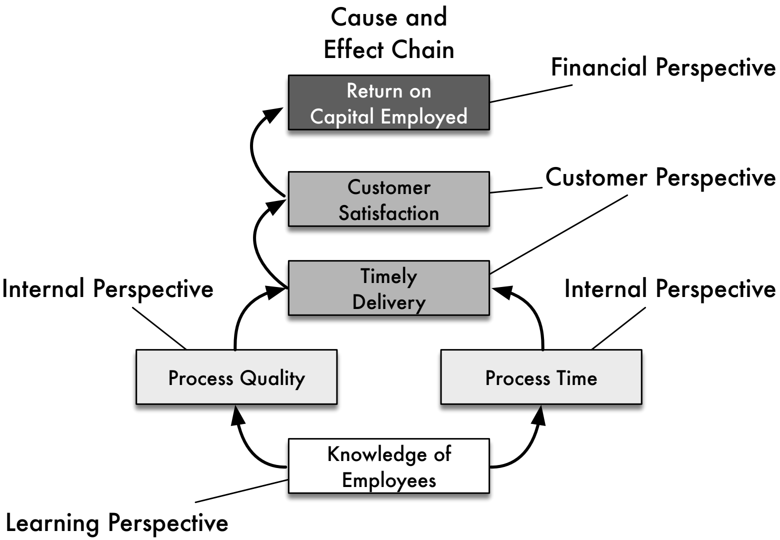
Reporting


Report
24.11.2011
Year 2011
Sales of RedWine: 12,2 Mio.
Sales of WhiteWine: 9,4 Mio.
Sales of RoseWine: 6,8 Mio.
1. Half
Sales of RedWine: 6,8 Mio.
Sales of WhiteWine: 6 Mio.
Sales of RoseWine: 2,5 Mio.
2.Half
Sales of RedWine: 5,4 Mio.
Sales of WhiteWine: 3,4 Mio.
Sales of RoseWine: 4,3 Mio.
WhiteWine
Rose
RedWine
Market Development
Turnover: 10,3 Mio
Revenue: 2,1 Mio
Production
Employee Satisfaction
Customer Satisfaction
Sales

Balanced Scorecard



Interdependence



Decision Support