

# Combining Data Mining and Decision Support

Marko Boharac

---

---

---

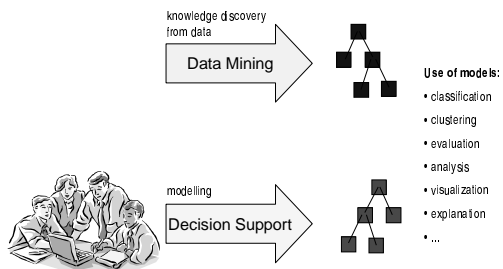
---

---

---

---

---



Marko Boharac

---

---

---

---

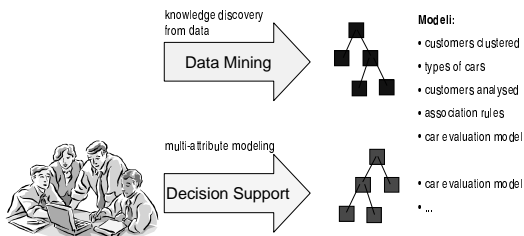
---

---

---

---

## Example: Cars



Marko Boharac

---

---

---

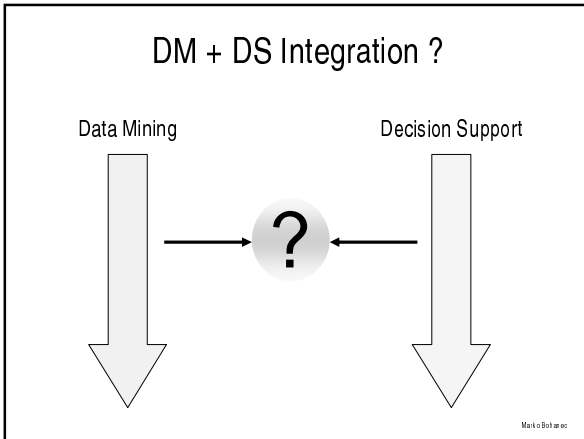
---

---

---

---

---




---

---

---

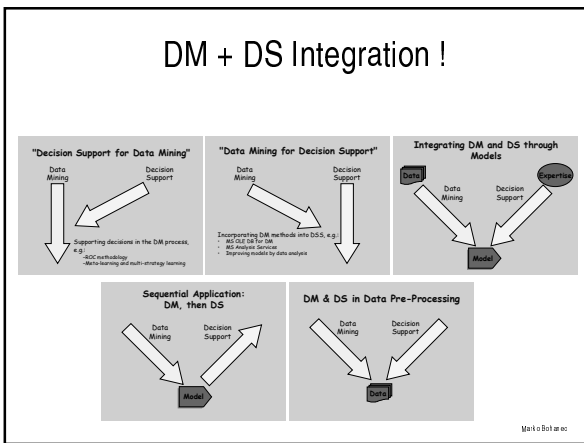
---

---

---

---

---




---

---

---

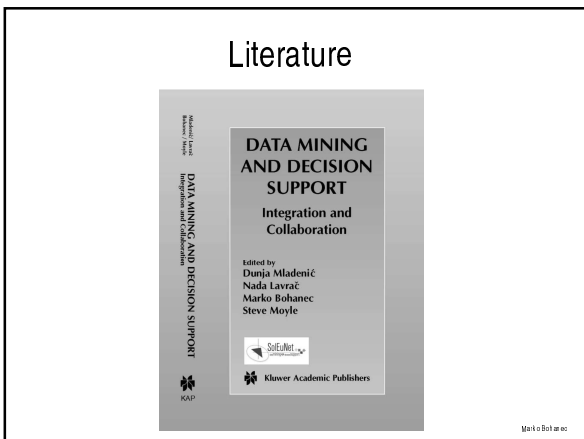
---

---

---

---

---




---

---

---

---

---

---

---

---

## Combining DM and DS

- “DS for DM”:
  - ROC methodology
  - meta-learning
- “DM for DS”:
  - MS OLE DB for DM
  - MS Analysis Services
  - model revision (from data)
- “DM, then DS” (sequential application):
  - Decisions-At-Hand approach
- “DS, then DM” (sequential application):
  - using models in data pre-processing for DM
- “DM and DS” (parallel application):
  - combining through models, e.g., DEXi in HINT
  - considering different problem dimensions

Marko Bohar et al.

---

---

---

---

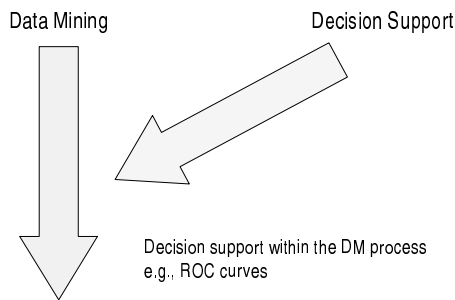
---

---

---

---

## “DS for DM”



Marko Bohar et al.

---

---

---

---

---

---

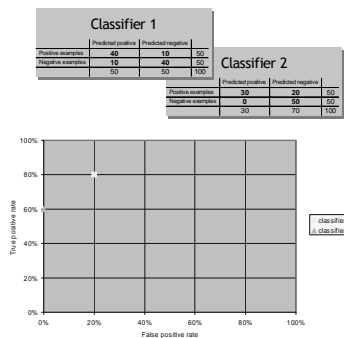
---

---

## ROC space

Peter Flach

- **True positive rate** =  
#true pos. / #pos.
  - $TPR_1 = 40/50 = 80\%$
  - $TPR_2 = 30/50 = 60\%$
- **False positive rate** =  
#false pos. / #neg.
  - $FPR_1 = 10/50 = 20\%$
  - $FPR_2 = 0/50 = 0\%$
- **ROC space** has
  - FPr on X axis
  - TPr on Y axis



Marko Bohar et al.

---

---

---

---

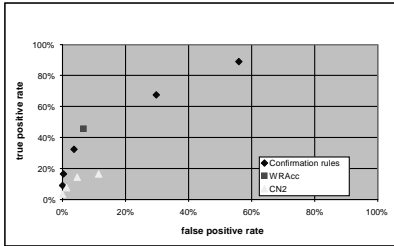
---

---

---

---

## The ROC convex hull



Marko Boharac

---

---

---

---

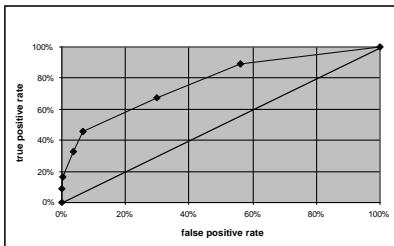
---

---

---

---

## The ROC convex hull



Marko Boharac

---

---

---

---

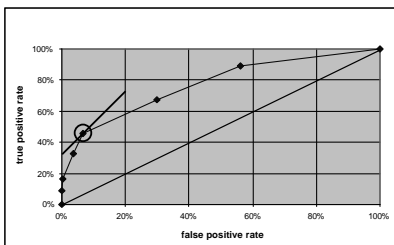
---

---

---

---

## Choosing a classifier



$$\frac{FP_{cost}}{FN_{cost}} = \frac{1}{2}$$

$$\frac{Neg}{Pos} = 4$$

$$slope = \frac{1}{2} = 2$$

Marko Boharac

---

---

---

---

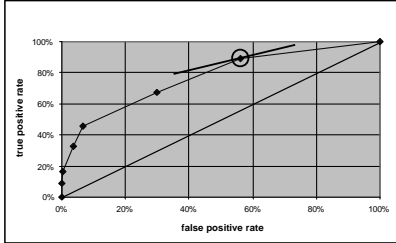
---

---

---

---

## Choosing a classifier



$$\frac{FPcost}{FNcost} = \frac{1}{8}$$

$$\frac{Neg}{Pos} = 4$$

$$slope = \frac{4}{8} = .5$$

Marko Boharac

---

---

---

---

---

---

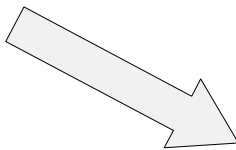
---

---

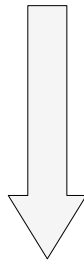
## "DM for DS"

Data Mining

Decision Support



Introducing DM methods into the DS process:  
 - MS SQL Server - Analysis Services  
 - model revision



Marko Boharac

---

---

---

---

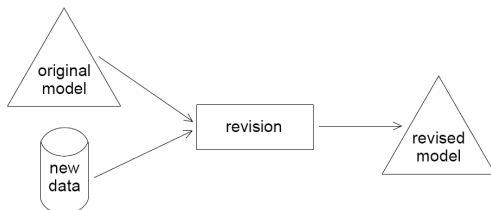
---

---

---

---

## "DM for DS": Model Revision



Marko Boharac

---

---

---

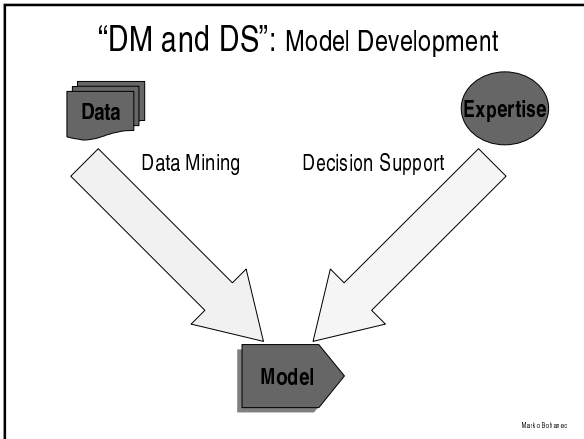
---

---

---

---

---




---

---

---

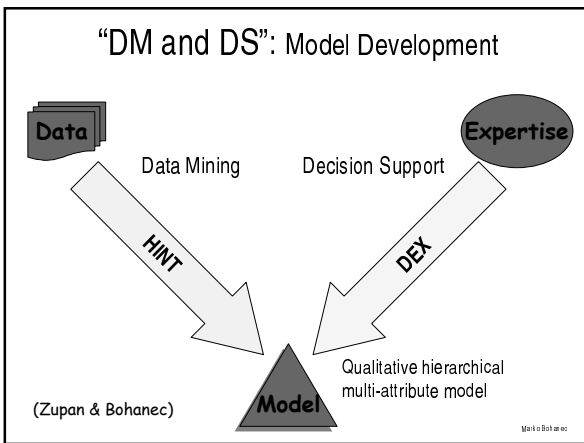
---

---

---

---

---




---

---

---

---

---

---

---

---

- ### Modes of Operation
1. **DEX only:** from expertise
  2. **HINT only:** from data
  3. **Supervised:** from data under expert supervision
  4. **Serial:** HINT-developed model subsequently refined by the expert
  5. **Parallel:** parallel development of model(s) by DEX and HINT
  6. **Combined:** combining sub-models developed in different ways
- Marko Bohanec

---

---

---

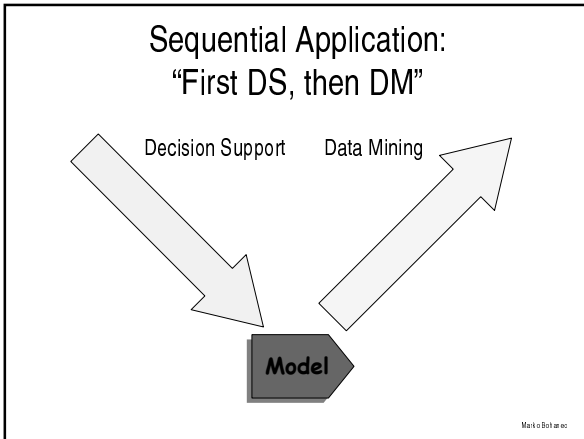
---

---

---

---

---




---

---

---

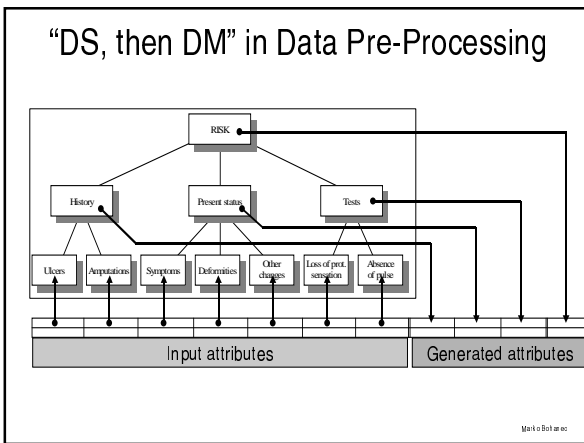
---

---

---

---

---




---

---

---

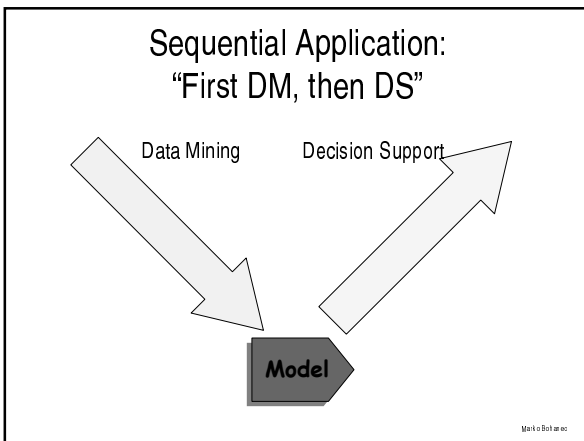
---

---

---

---

---




---

---

---

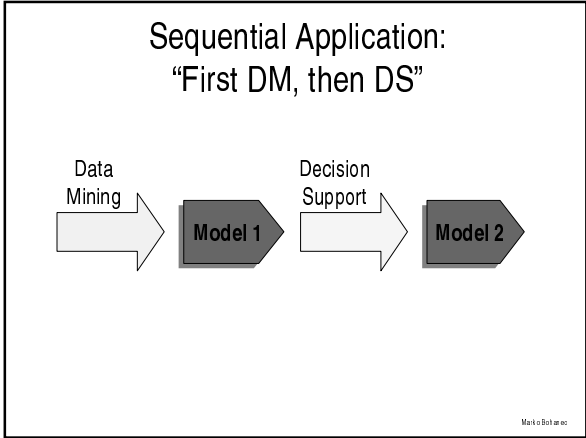
---

---

---

---

---




---

---

---

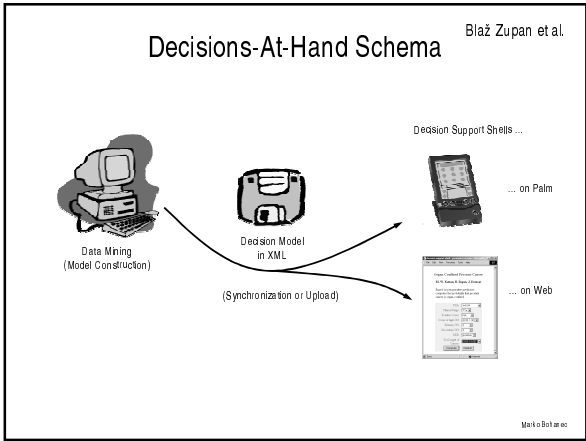
---

---

---

---

---




---

---

---

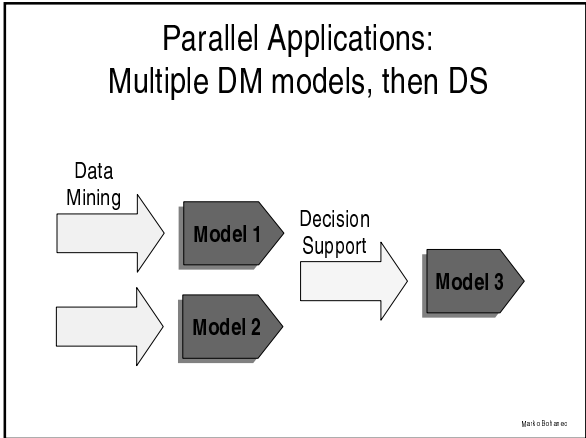
---

---

---

---

---




---

---

---

---

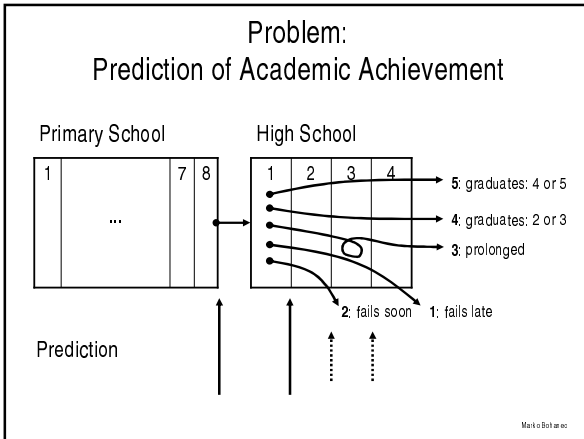
---

---

---

---






---

---

---

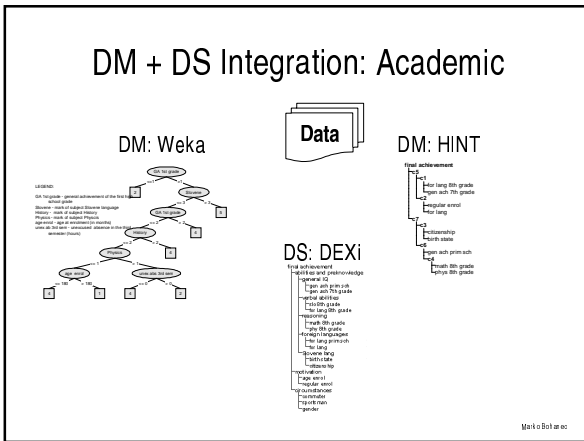
---

---

---

---

---




---

---

---

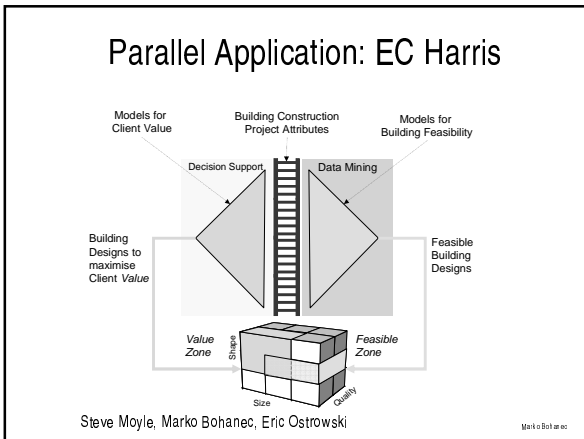
---

---

---

---

---




---

---

---

---

---

---

---

---

## Conclusion

- DM & DS approaches are:
  - complementary
  - supplementary
- New and developing research area
- Typical combinations:
  - *DS for DM*
  - *DM for DS*
  - *DM, then DS*
  - *DS, then DM*
  - *DM and DS*
- Open questions:
  - formalization (framework) of DM&DS integration
  - common methodologies and approaches
  - standardization

Marko Böhmer

---

---

---

---

---

---

---

---