Decision support systems remade: (machine) learning advisors

Filip Wójcik
Data Scientist
Wroclaw University Lecturer

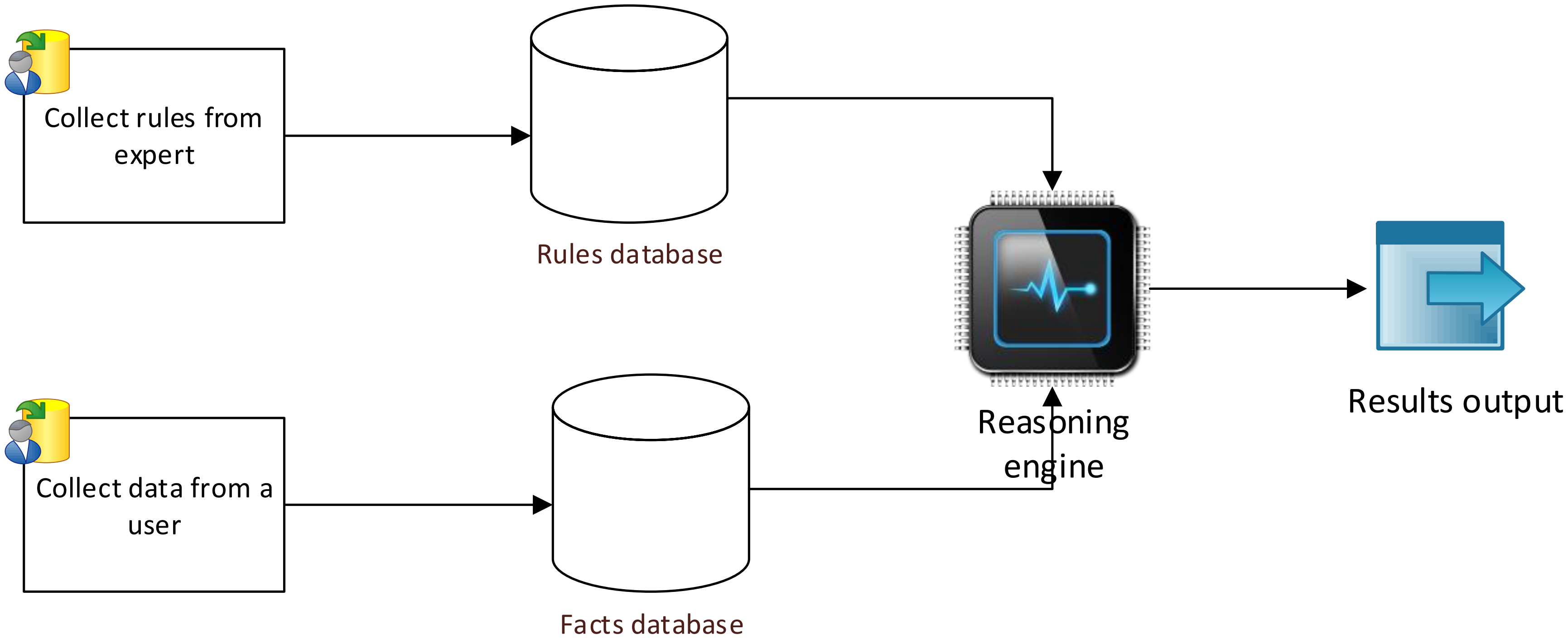
Agenda

- What are decision support and expert systems?
- Brief history of expert systems
- Drawbacks of traditional expert systems
- Machine learning – modern decision support systems
- Association analysis
- Rule induction
- Decision trees
- Scalable solutions

What are expert systems? 1/4

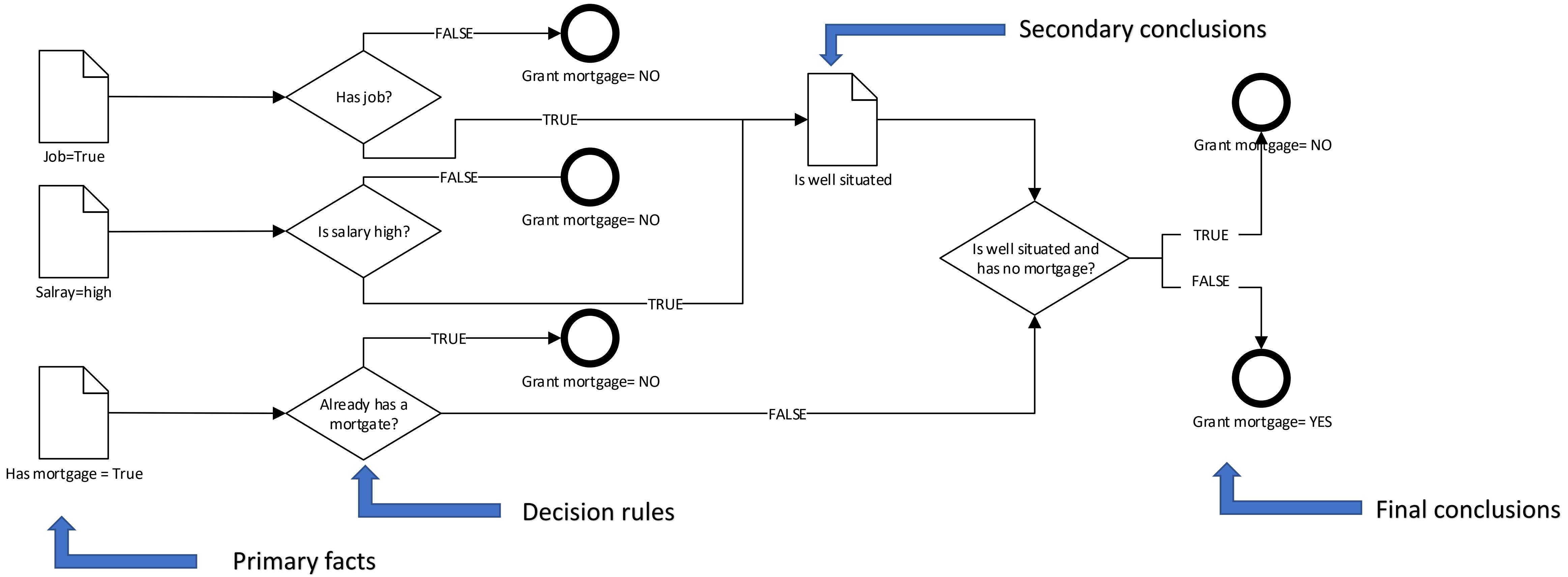
- Storing experts knowledge
- Reasoning over certain facts
- Quick reactions on user queries
- Transparency of decisions
- Explainability of decisions

What are expert systems? 2/4

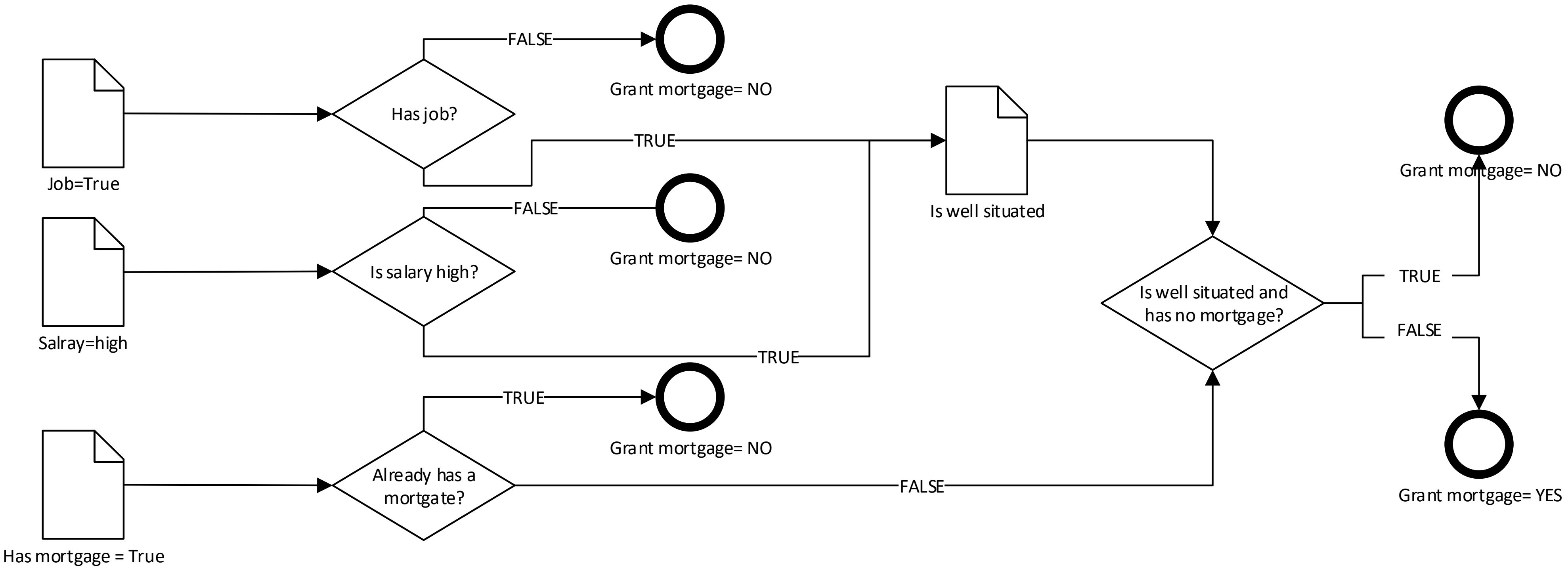


1. Collect rules from domain-experts
2. Collect facts from users
3. Perform reasoning over facts, using rules
4. Conclude facts
5. Present results of reasoning

What are expert systems? 3/4



What are expert systems? 4/4



Reasoning FORWARD: from facts to conclusions

Reasoning BACKWARD: from conclusions to facts

Brief history of expert systems

Early 70's

- First trials
- Attempt to mimic experts knowledge in computers

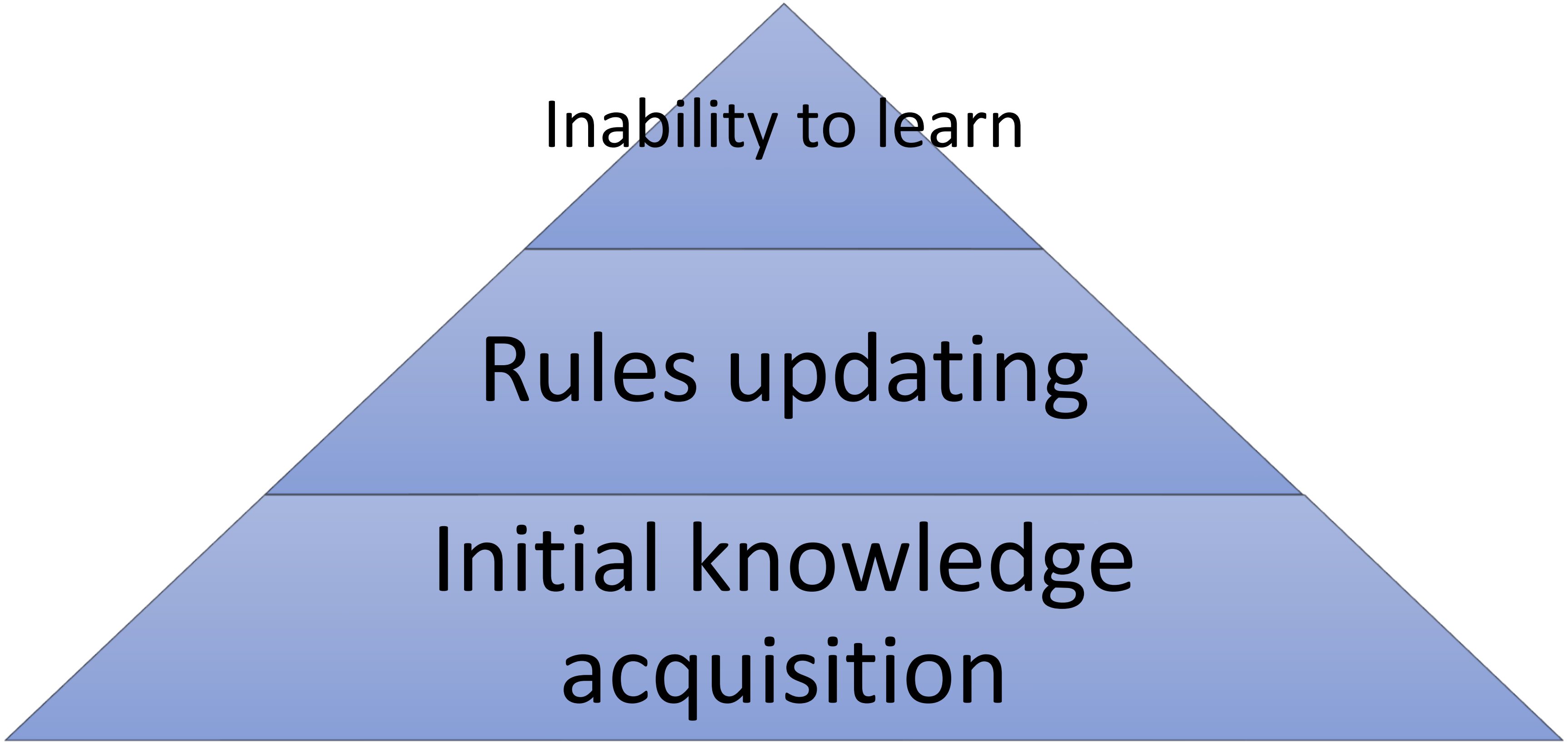
80's

- IBM's research
- Expert systems usage in medicine, mining,minerals searching

90's

DEATH

What killed expert systems?

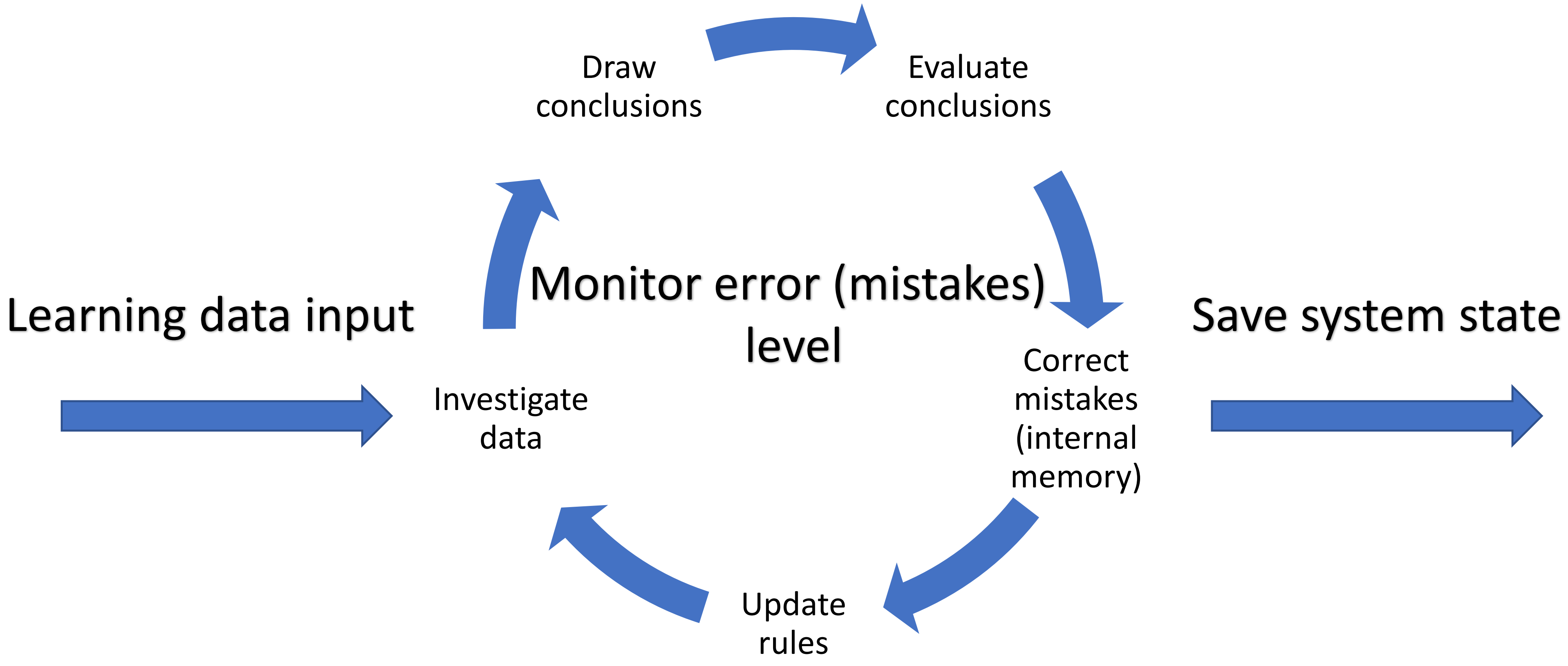


Machine learning – a modern decision support systems

Machine learning – a modern decision suport systems 1/3

- Ability to automatically learn from data
- Usage of statistical procedures to validate knowledge
- Variety of tasks
- Explainability of decisions (sometimes)
- Easy to validate/score

Machine learning – a modern decision suport systems 2/3



Machine learning – a modern decision suport systems 3/3

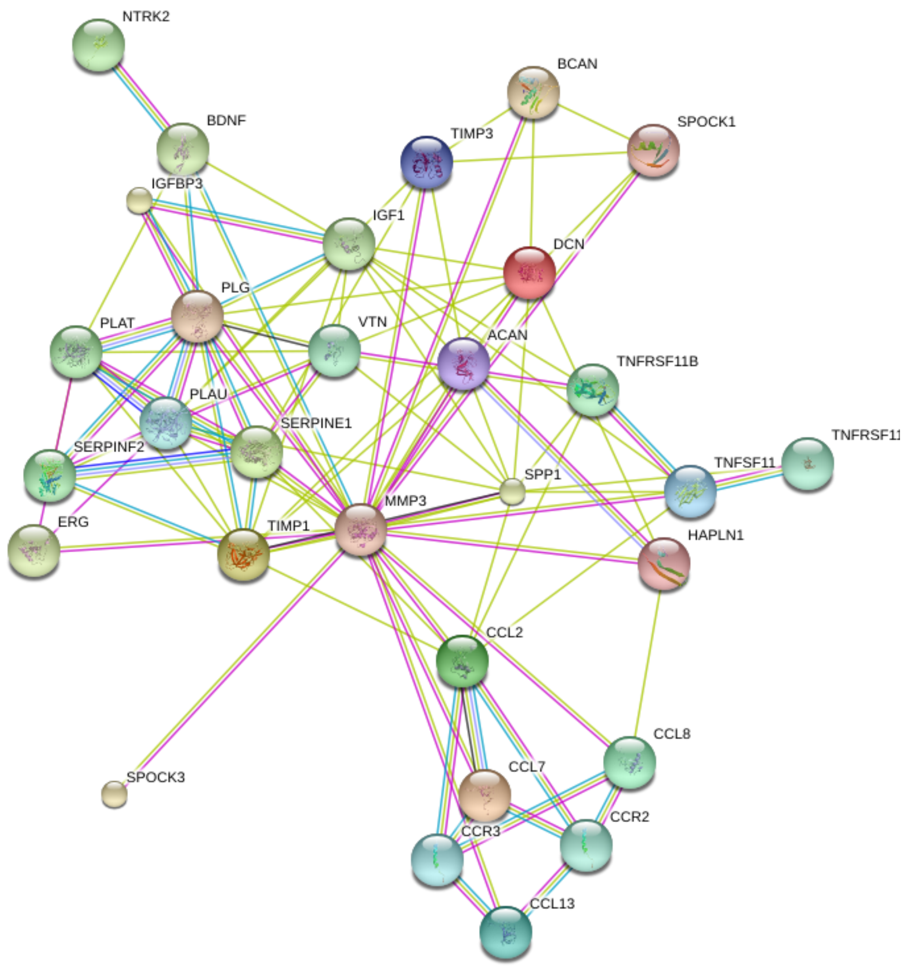
- Cannot be interpreter by humans
- Their internal structure is complicated and is hard to understand
- „Justifications” of predictions are purely mathematical

„Black box” methods

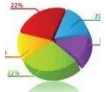
- Easily interpretable
- Can be translated to human-friendly form
- Not so sophisticated mathematically

„White box” methods

Association analysis systems



Association analysis systems 1/6

Key features	Use cases
Easy to interpret	 Market basket analysis
 Wide range of use cases	Social media graph structures analysis
 Dual role: explaining data/making decisions	 Correlation fining
Scalable	 Consumer profiling

Association analysis systems 2/6

Transaction number	Products
1.	1. Soya milk 2. Salad
2.	1. Salad 2. Walnuts 3. Wine 4. Bread
3.	1. Soya milk 2. Walnuts 3. Wine 4. Juice
4.	1. Salad 2. Soya milk 3. Walnuts 4. Wine
5.	1. Salad 2. Soya milk 3. Walnuts 4. Juice

Normalized vs denormalized form

Order Id	Product category	city	state
1006414	BOOK	BURMINGHAM	AL
1006541	ARTWORK	MADISON	CT
1006542	OCCASION	OLD SAYBROOK	CT
1010154	BOOK	EASTCHESTER	NY
1009110	ARTWORK	LINCON PARL	NJ
1009110	BOOK	LINCON PARL	NJ
1008491	ARTWORK	TARRYTOWN	NY
1008492	ARTWORK	CARMEL	NY
1010189	BOOK	WHITE PLAINS	NY
1008493	BOOK	WHITE PLAINS	NY

Association analysis systems 3/6

Transaction number	Products
1.	1. Soya milk 2. Salad
2.	1. Salad 2. Walnuts 3. Wine 4. Bread
3.	1. Soya milk 2. Walnuts 3. Wine 4. Juice
4.	1. Salad 2. Soya milk 3. Walnuts 4. Wine
5.	1. Salad 2. Soya milk 3. Walnuts 4. Juice

IF

THEN

Antecedent

Consequent

Frequent items	Rel. support
Soya, salad	0.6
Soya, salad, walnuts	0.4
Salad	0.8

Implications	confidence
Soya => walnuts	0.75
Soya => salad	0.75
Soya, Walnuts, Wine => juice	0.4

Key features:

- Finding items that co-occur together
- Finding items that are correlated
- Form asiciation rules, defining relationships
- Calculate probabilities and confidence values

Association analysis systems 4/6

Transaction number	Products
1.	1. Soya milk 2. Salad
2.	1. Salad 2. Walnuts 3. Wine 4. Bread
3.	1. Soya milk 2. Walnuts 3. Wine 4. Juice
4.	1. Salad 2. Soya milk 3. Walnuts 4. Wine
5.	1. Salad 2. Soya milk 3. Walnuts 4. Juice

Frequent items	Rel. support
Soya, salad	0.6
Soya, salad, walnuts	0.4
Salad	0.8

Implications	confidence
walnuts => soya	0.75
Soya => salad	0.75
Soya, Walnuts, Wine => juice	0.4

$$support(X) = count(X)$$

$$relative\ support(X\&Y) = \frac{support(X\ \&\ Y)}{N} = P(AB)$$

$$confidence(X\ \rightarrow Y) = \frac{support(X\ \&\ Y)}{support(X)} = P(A|B)$$

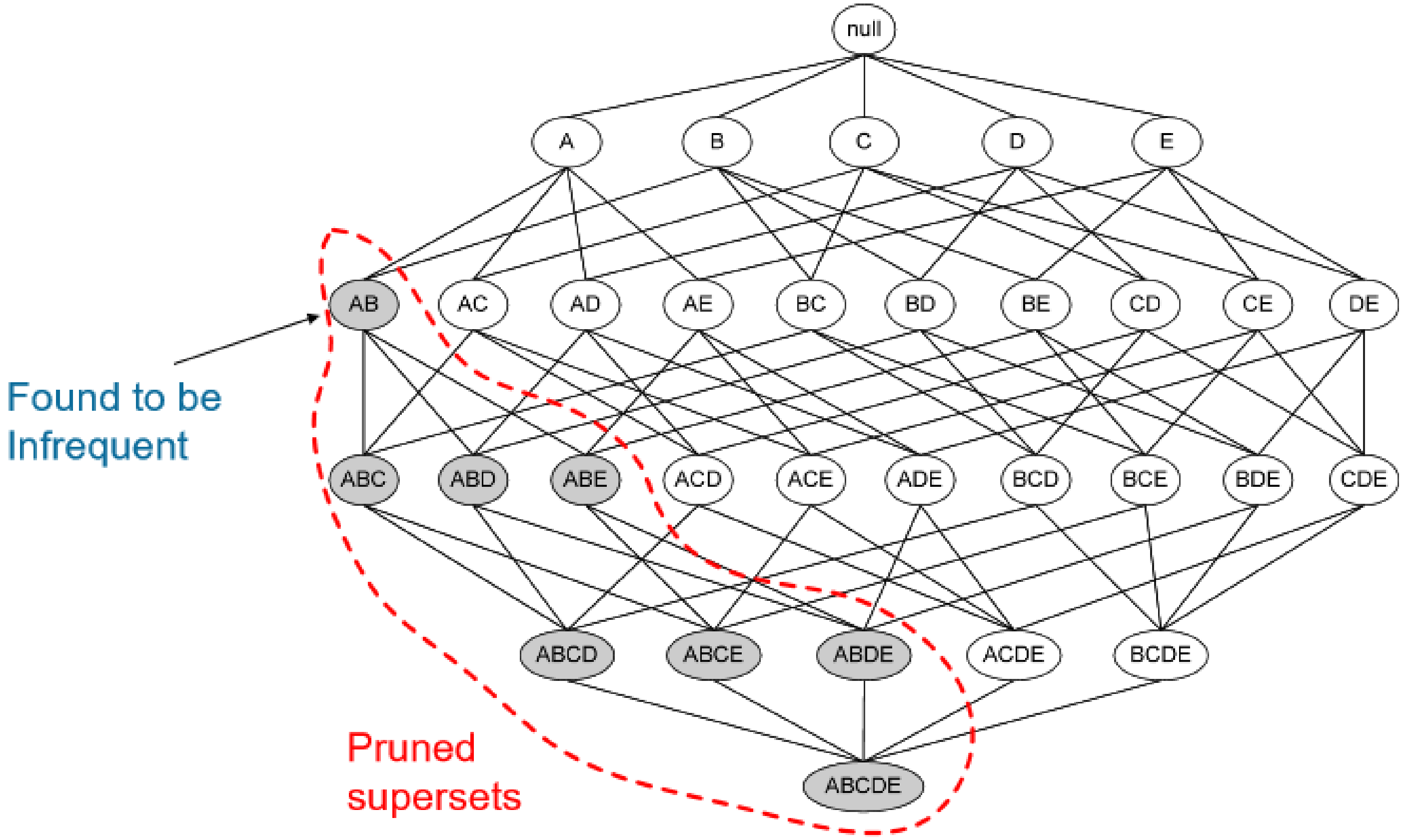
$$relative\ support(soya\ \&salad) = \frac{3}{5} = 0.6$$

$$confidence(soya\ \rightarrow salad) = \frac{3}{4} = 0.75$$

Association analysis systems 5/6

#	Measure	Formula
1	ϕ-coefficient	$\frac{P(A,B)-P(A)P(B)}{\sqrt{P(A)P(B)(1-P(A))(1-P(B))}}$
2	Goodman-Kruskal's (λ)	$\frac{\sum_j \max_k P(A_j,B_k)+\sum_k \max_j P(A_j,B_k)-\max_j P(A_j)-max_k P(B_k)}{2-\max_j P(A_j)-\max_k P(B_k)}$
3	Odds ratio (α)	$\frac{P(A,B)P(\overline{A},\overline{B})}{P(A,\overline{B})P(\overline{A},B)}$
4	Yule's Q	$\frac{P(A,B)P(\overline{AB})-P(A,\overline{B})P(\overline{A},B)}{P(A,B)P(\overline{AB})+P(A,\overline{B})P(\overline{A},B)} = \frac{\alpha-1}{\alpha+1}$
5	Yule's Y	$\frac{\sqrt{P(A,B)P(\overline{AB})}-\sqrt{P(A,\overline{B})P(\overline{A},B)}}{\sqrt{P(A,B)P(\overline{AB})}+\sqrt{P(A,\overline{B})P(\overline{A},B)}} = \frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}$
6	Kappa (κ)	$\frac{P(A,B)+P(\overline{A},\overline{B})-P(A)P(B)-P(\overline{A})P(\overline{B})}{1-P(A)P(B)-P(\overline{A})P(\overline{B})}$
7	Mutual Information (M)	$\frac{\sum_i \sum_j P(A_i,B_j) \log \frac{P(A_i,B_j)}{P(A_i)P(B_j)}}{\min(-\sum_i P(A_i) \log P(A_i), -\sum_j P(B_j) \log P(B_j))}$
8	J-Measure (J)	$\max\big(P(A,B)\log(\frac{P(B\|A)}{P(B)}) + P(A\overline{B})\log(\frac{P(\overline{B}\|A)}{P(\overline{B})}),$ $P(A,B)\log(\frac{P(A\|B)}{P(A)}) + P(\overline{A}B)\log(\frac{P(\overline{A}\|B)}{P(\overline{A})})\big)$
9	Gini index (G)	$\max\big(P(A)[P(B\|A)^2 + P(\overline{B}\|A)^2] + P(\overline{A})[P(B\|\overline{A})^2 + P(\overline{B}\|\overline{A})^2]$ $-P(B)^2 - P(\overline{B})^2,$ $P(B)[P(A\|B)^2 + P(\overline{A}\|B)^2] + P(\overline{B})[P(A\|\overline{B})^2 + P(\overline{A}\|\overline{B})^2]$ $-P(A)^2 - P(\overline{A})^2\big)$
10	Support (s)	$P(A,B)$
11	Confidence (c)	$\max(P(B\|A), P(A\|B))$
12	Laplace (L)	$\max\big(\frac{NP(A,B)+1}{NP(A)+2}, \frac{NP(A,B)+1}{NP(B)+2}\big)$
13	Conviction (V)	$\max\big(\frac{P(A)P(\overline{B})}{P(A\overline{B})}, \frac{P(B)P(\overline{A})}{P(B\overline{A})}\big)$
14	Interest (I)	$\frac{P(A,B)}{P(A)P(B)}$
15	cosine (IS)	$\frac{P(A,B)}{\sqrt{P(A)P(B)}}$
16	Piatetsky-Shapiro's (PS)	$P(A,B) - P(A)P(B)$
17	Certainty factor (F)	$\max\big(\frac{P(B\|A)-P(B)}{1-P(B)}, \frac{P(A\|B)-P(A)}{1-P(A)}\big)$
18	Added Value (AV)	$\max(P(B\|A) - P(B), P(A\|B) - P(A))$
19	Collective strength (S)	$\frac{P(A,B)+P(\overline{AB})}{P(A)P(B)+P(\overline{A})P(\overline{B})} \times \frac{1-P(A)P(B)-P(\overline{A})P(\overline{B})}{1-P(A,B)-P(\overline{AB})}$
20	Jaccard (ζ)	$\frac{P(A,B)}{P(A)+P(B)-P(A,B)}$
21	Klosgen (K)	$\sqrt{P(A,B)}\max(P(B\|A) - P(B), P(A\|B) - P(A))$

Association analysis systems 6/6



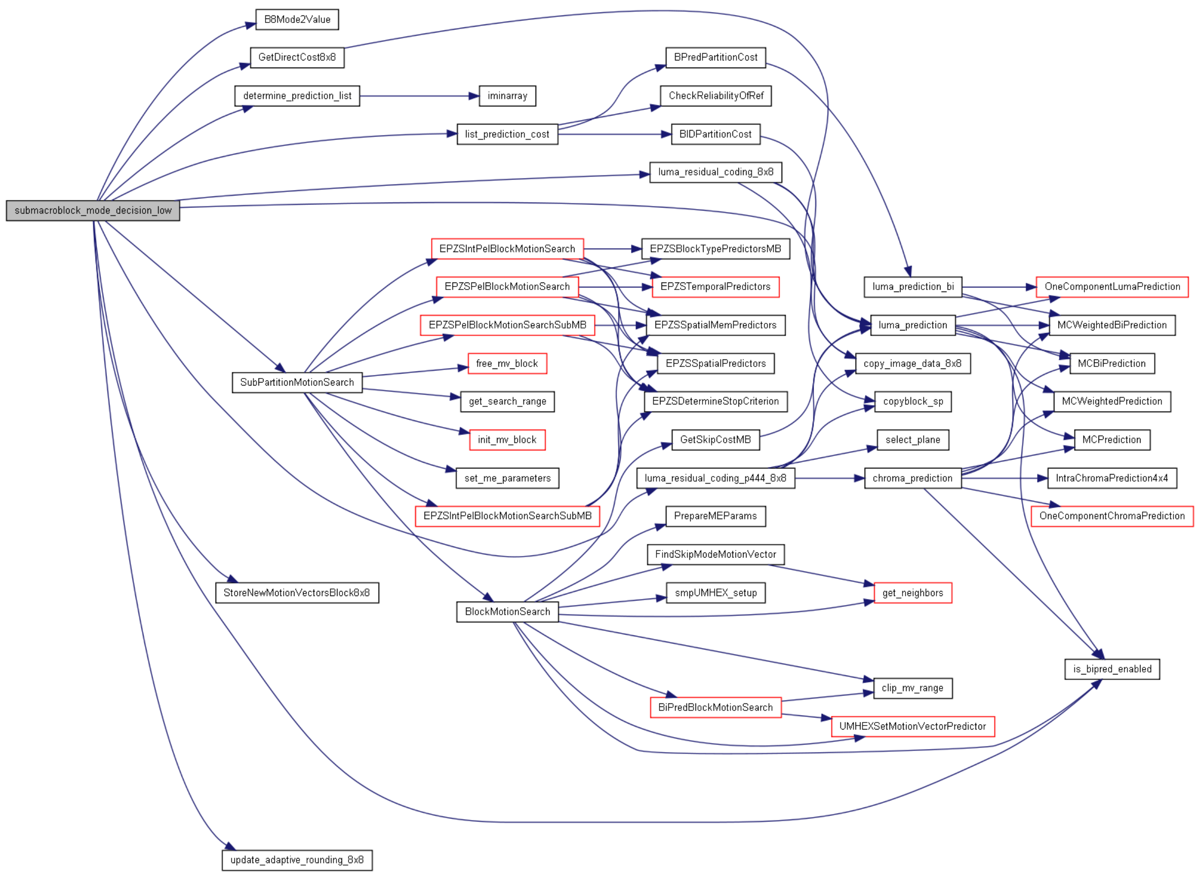
Apriori principle:

- Given user-defined frequency threshold **T**
- Given infrequent item **Itm (suport(Itm) < T)**
- No itemset containing **Itm can be frequent**

Result:

- Less combinatorial effort
- No need to check infrequent combinations

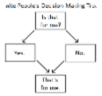
Rule induction systems



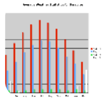
Rule induction systems 1/5

Key features

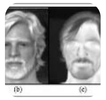

Full decision transparency

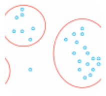

Full decision explainability


Automatic knowledge updating


Based on statistical properties of data

Use cases


Classification systems


Pattern recognition


Intelligent advisors

Rule induction systems 2/5

client	hotel	addons	money_spent	offer
business	Hilton	trip	40,000	deluxe
business	Hilton	full board	38,000	deluxe
business	Hilton	trip	40,000	deluxe
middle class	Meta	none	800	basic
middle class	Meta	trip	900	basic
manager	Meta	trip	1,500	premium

Value	Count	%
Deluxe	3	0.5
Basic	2	0.333
Premium	1	0.16666

Decisive attribute

Rule induction systems 3/5

client	hotel	addons	money_spent	offer
business	Hilton	trip	40,000	deluxe
business	Hilton	full board	38,000	deluxe
business	Hilton	trip	40,000	deluxe
middle class	Meta	none	800	basic
middle class	Meta	trip	900	basic
manager	Meta	trip	1,500	premium

IF hotel == „Meta” **THEN** offer **==**

Value	Count	Perc.
basic	2	66.6
premium	1	33.3

IF addons == „trip” **THEN** offer **==**

Value	Count	Perc.
deluxe	2	50
basic	1	25
premium	1	25

Rule induction systems 4/5

client	hotel	addons	money_spent	offer
business	Hilton	trip	40,000	deluxe
business	Hilton	full board	38,000	deluxe
business	Hilton	trip	40,000	deluxe
middle class	Meta	none	800	basic
middle class	Meta	trip	900	basic
manager	Meta	trip	1,500	premium

IF

hotel == „Meta” **AND**
client == „middle class”

THEN

Value	Count	Perc.
basic	2	100

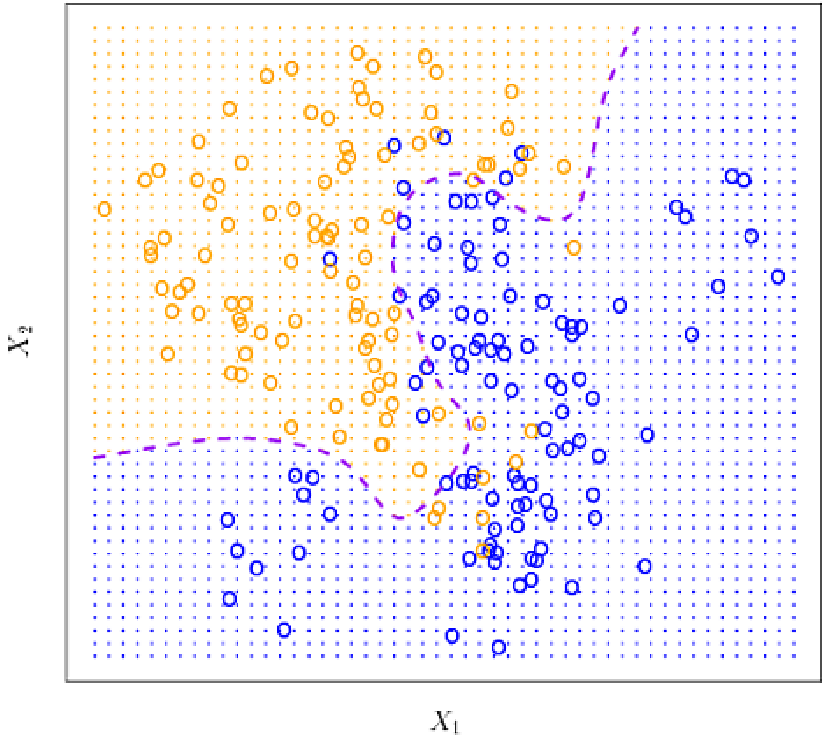
IF

hotel == „Meta” **AND**
client == „manager”

THEN

Value	Count	Perc.
premium	1	100

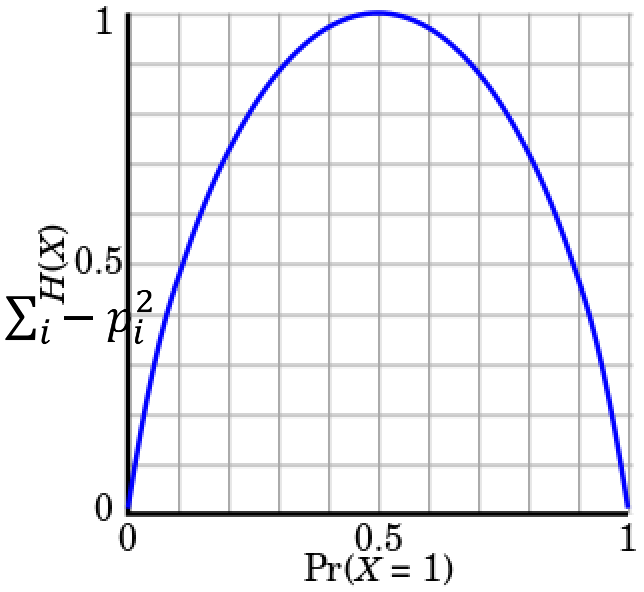
Rule induction systems 5/5



Key features:

- Incremental building of decision boundary – separating classes
- Adding conditions to rule antecedent to find a best separation between classes
- Several statistical measures:
    - Shannon entropy

$$Entropy = \sum_i - p_i \ln p_i$$



    - Gini impurity

$$Gini\ impiruty = 1 - \sum_i - p_i^2$$

    - Rule accuracy

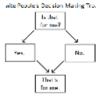
$$coverage = \frac{|records\ meeting\ antecedent\ \&\ consequent|}{|records\ meeting\ antecedent|}$$

Decision trees

Decision trees 1/5

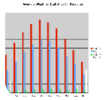
Key features

Full decision transparency



Full decision explainability

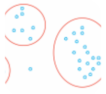
Compact data representation



Based on statistical properties of data

Use cases

Classification systems



Pattern recognition

Intelligent advisors

Visualization capabilities

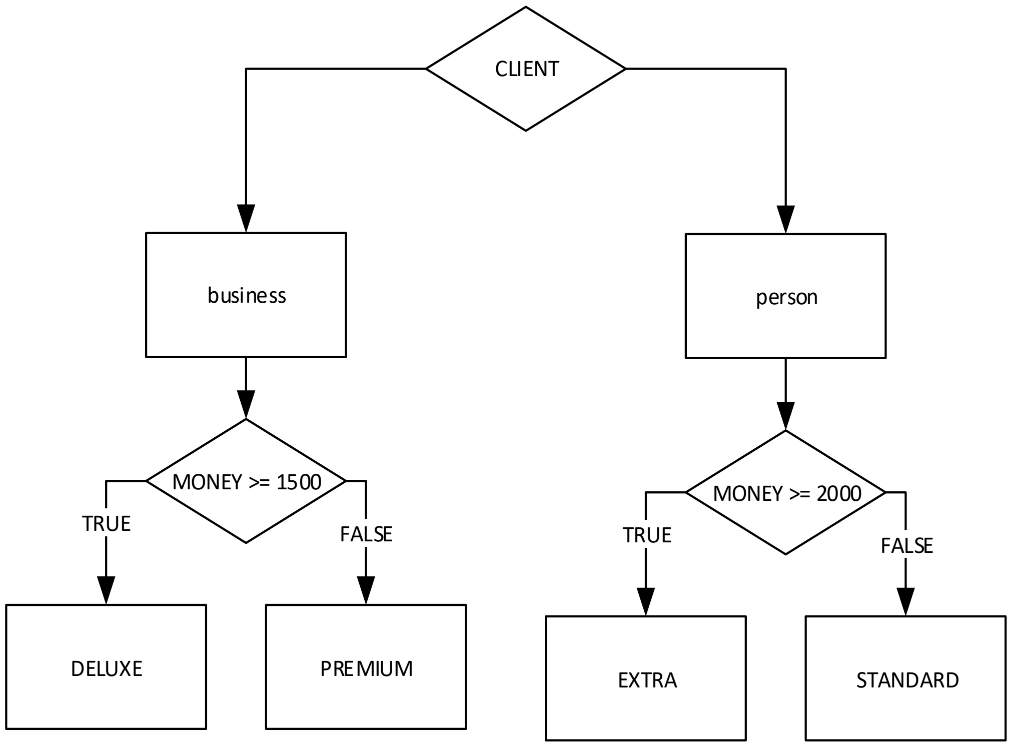
Decision trees 2/5

Client == business + Hotel == Hilton + Money >= 1500 ➜ DELUXE

Client == business + Hotel == Hilton + Money < 1500 ➜ PREMIUM

Client == person + Hotel == Art + Money >= 2000 ➜ EXTRA

Client == person + Hotel == Art + Money < 2000 ➜ STANDARD



Decision trees 3/5

client	hotel	addons	money_spent	offer
business	Hilton	trip	40,000	deluxe
business	Hilton	full board	38,000	deluxe
business	Hilton	trip	40,000	deluxe
middle class	Meta	none	800	basic
middle class	Meta	meal	900	basic
manager	Meta	spa	1,500	premium

Client == business?

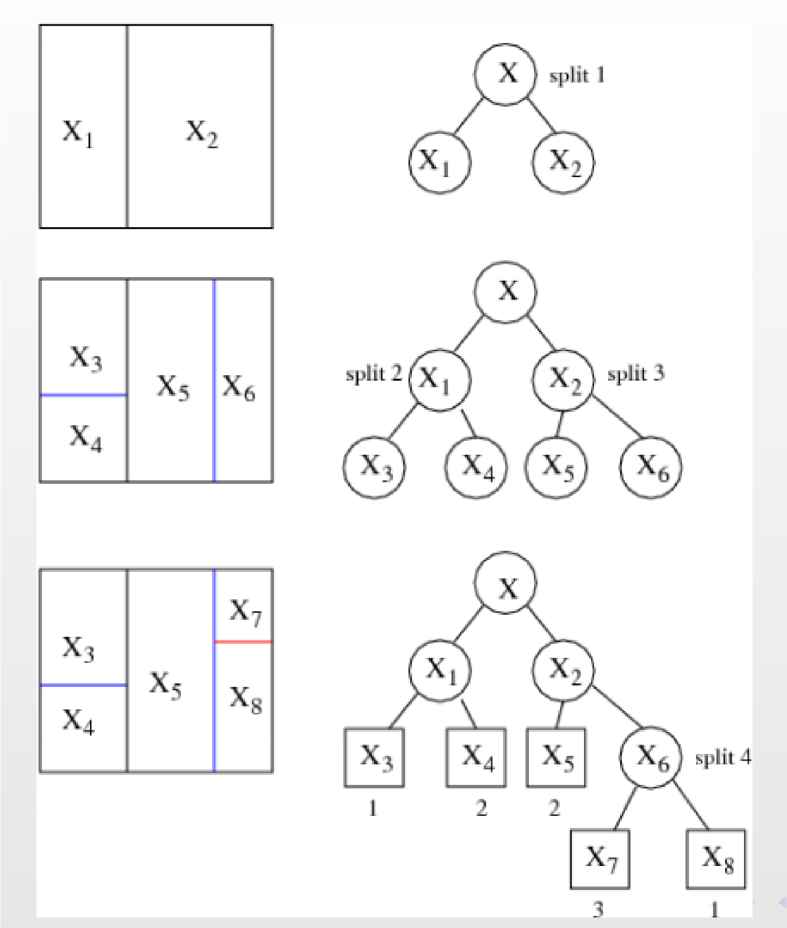
True

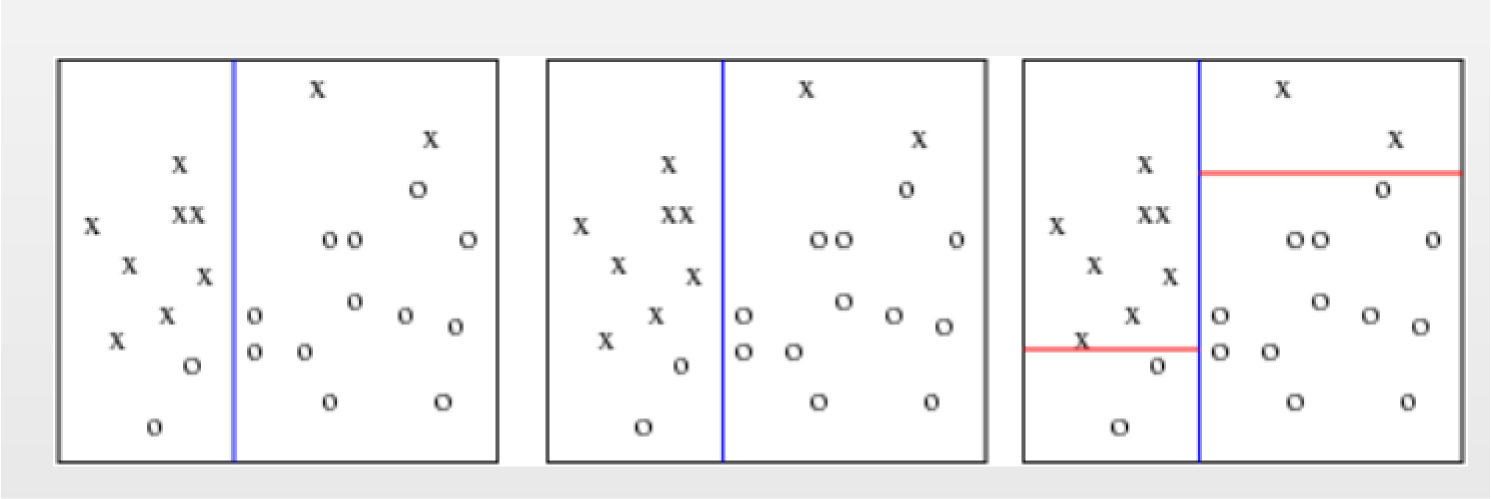
hotel	addons	money_spent	offer
Hilton	trip	40,000	deluxe
Hilton	full board	38,000	deluxe
Hilton	trip	40,000	deluxe

False

hotel	addons	money_spent	offer
Meta	none	800	basic
Meta	meal	900	basic
Meta	spa	1,500	premium

Decision trees 4/5





Decision trees 5/5

Recursive

- At each level performs the same operation
- Top – down approach
- Divide & conquer in construction
- Log(N) in traversal

Greedy

- Tries to reduce „chaos” at each split
- $Entropy = \sum_i -p_i \ln p_i$
- Trying to impose ORDER on data

Two forms

- Binary
  - Yes/no splits
  - Single attribute value
- Multi-way
  - Split on every value
  - Multi-value splits

Scalable solutions

Scalable solutions 1/2

Complexity

- ML algorithms are very complex
- Traversing datasets multiple Times
- Combinations/Correlations/etc.

Data storage

- ML algorithms require access to full datasets
- Big volumes of data don’t fit into memory

Extensibility

- Adjusting algorithm configuration for a specific needs
- Customization of processing methods

Scalable solutions 2/2

Thank you for attention!