Phantom Works > Engineering and Information Technology
Mathematics and Computing Technology

Visual Data Mining

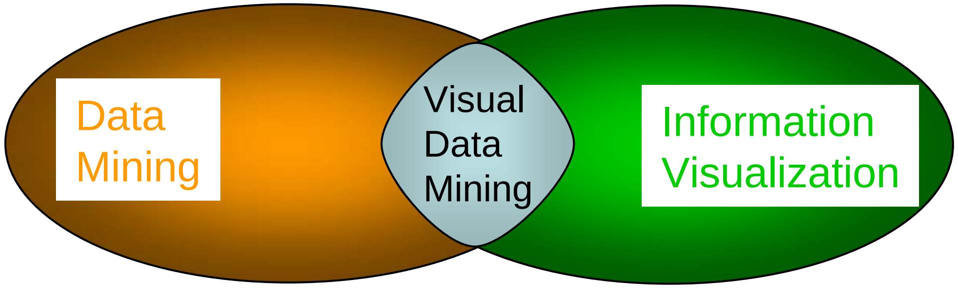
SC4Devo – July 15th 2004

Mihael Ankerst

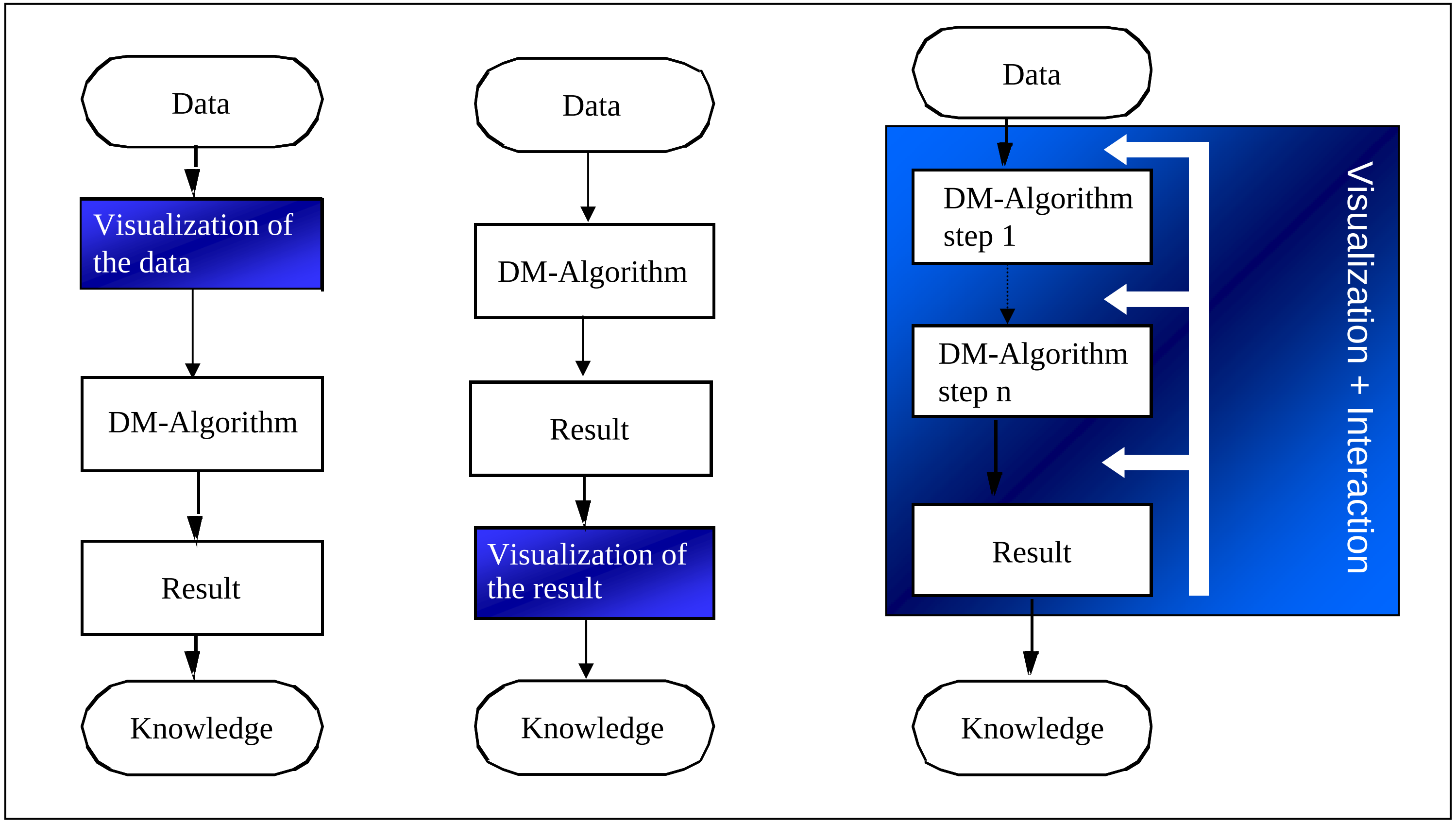
Boeing Phantom Works

Visual Data Mining

	Data Mining Algorithms	Visualization
Actionable	+	-
Evaluation	+	-
Flexibility	-	+
User Interaction	-	+



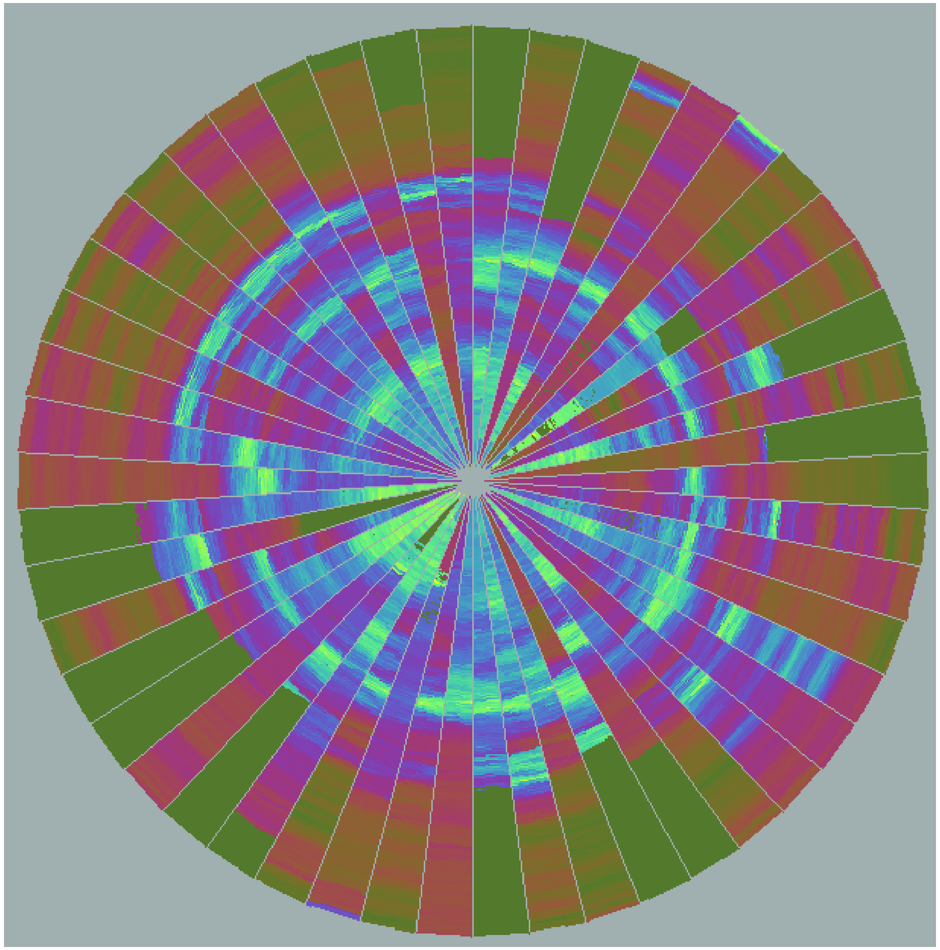
Visual Data Mining Architecture: Tightly Integrated Visualization



One example for <u>Preceding Visualization</u>

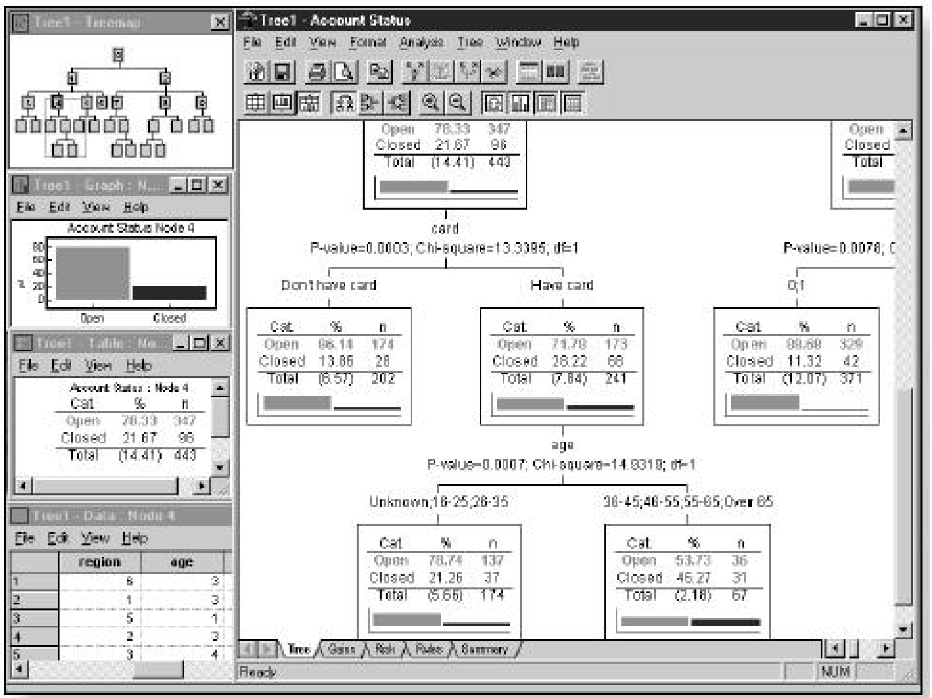
Circle Segments
Visualization of Stock Data

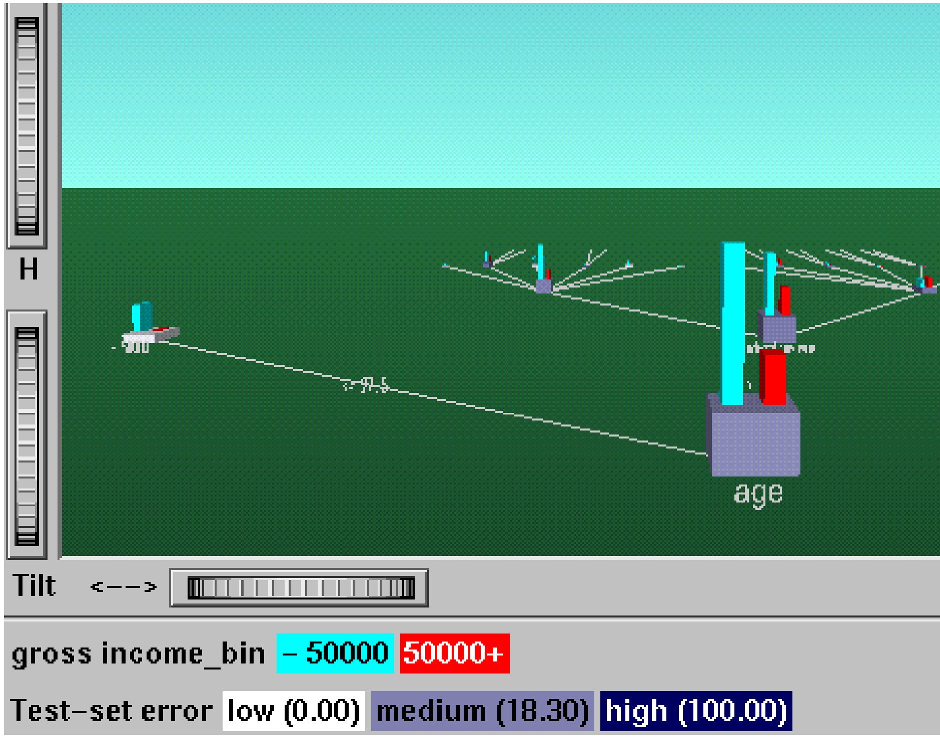
- ❑ Exploring ~10,000 records
50 different stock prices



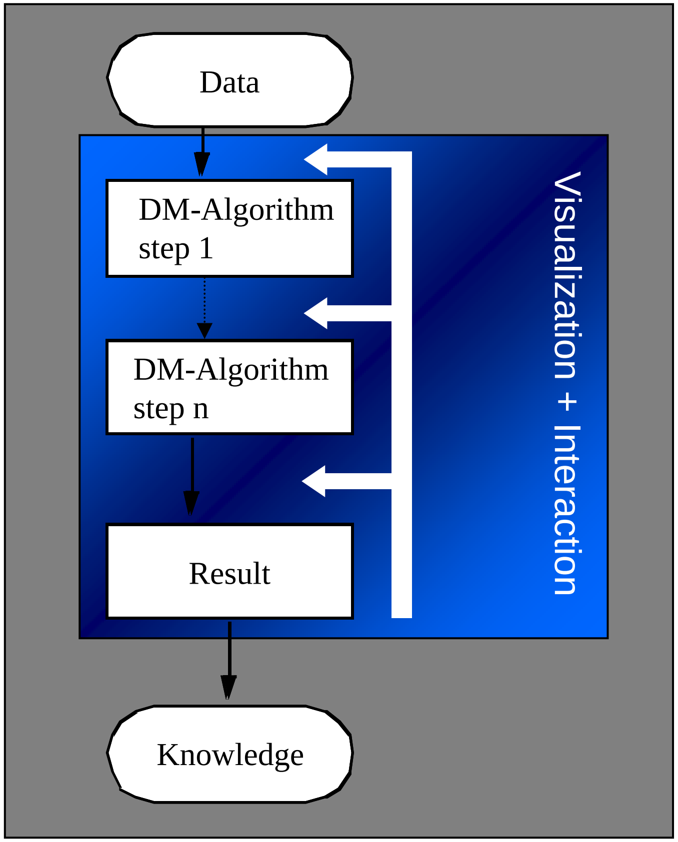
Two examples for Subsequent Visualization

- SPSS AnswerTree
- Decision Tree Visualizer (MineSet)





Tightly Integrated Visualization



Visualization of algorithmic decisions

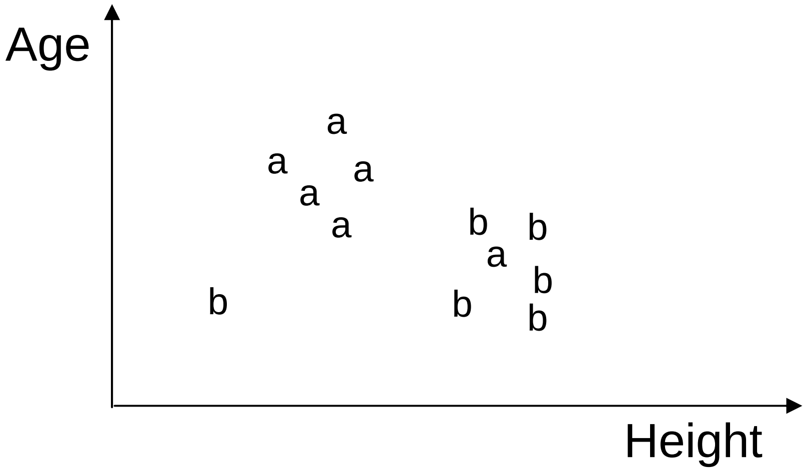
- ❑ Data and patterns are better understood
- ❑ User can make decisions based on perception
- ❑ User can make decisions based on domain knowledge
- ❑ Visualization of result enables user specified feedback for next algorithmic run

Tightly Integrated Visualization

- The first prototypes which follow this architecture:
  - Perception-Based Classification (Decision Tree Classification)
  - HD-Eye (Clustering)
  - DataJewel (Temporal Mining)

The corresponding DM Method

Classification



Problem description:

Given a set of objects with known class labels.

- Description

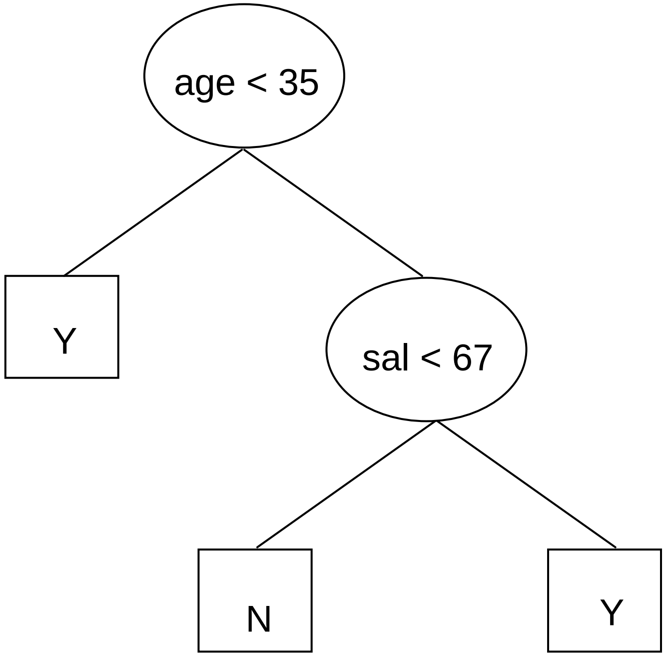
  Build model describing the data with respect to the class

- Prediction

  Use model to predict the class label of objects

Tree Model Tutorial

Age	Salary	Sex	**Class**
25	15	M	**Y**
42	40	M	**N**
29	63	F	**Y**
81	45	F	**N**
57	89	M	**Y**



<u>Problem description</u>:

Given data describing individuals, find factors that indicate buyers.

- Algorithm

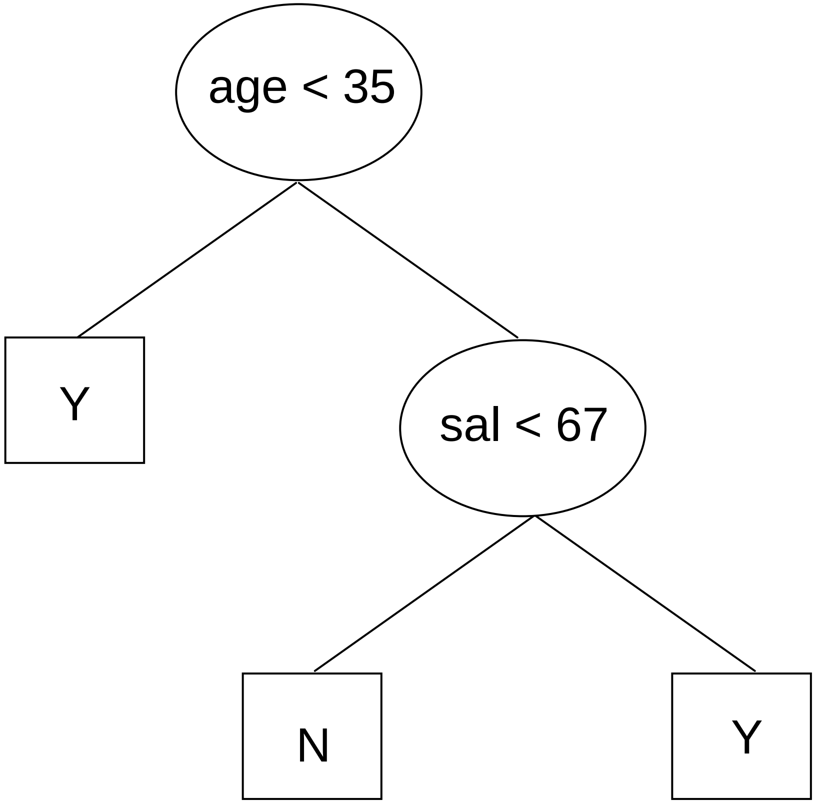
  Search through all factors to find one which best divides people into buyers / not buyers

  Divide groups and repeat on subgroups

- Outcome

  Tree uses factors to describe people who are likely to be buyers

Tutorial (2)
Extracting Rules from a Decision Tree



IF (age < 35) THEN Class = 'Yes'

IF (age >= 35) AND (sal < 67)

THEN Class = 'No'

IF (age >= 35) AND (sal >= 67)

THEN Class = 'Yes'

Visual Classification

attr.1	attr.2	...	class
0.3	23.3	**...**	**Y**
2.4	2.0	**...**	**N**
⋮	⋮	⋮	⋮

→

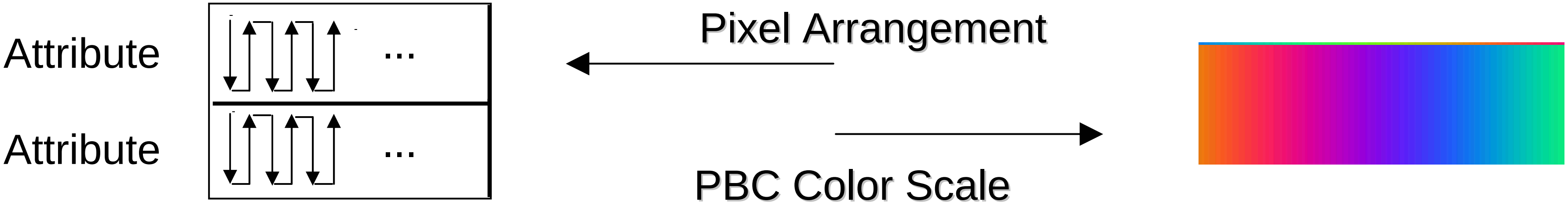
attr. 1	class
0.2	**Y**
0.3	**Y**
0.3	**Y**
0.5	**N**
1.1	**Y**
⋮	⋮

attr. 2	class
0.5	**N**
1.3	**Y**
2.0	**N**
2.5	**Y**
5.1	**N**
⋮	⋮

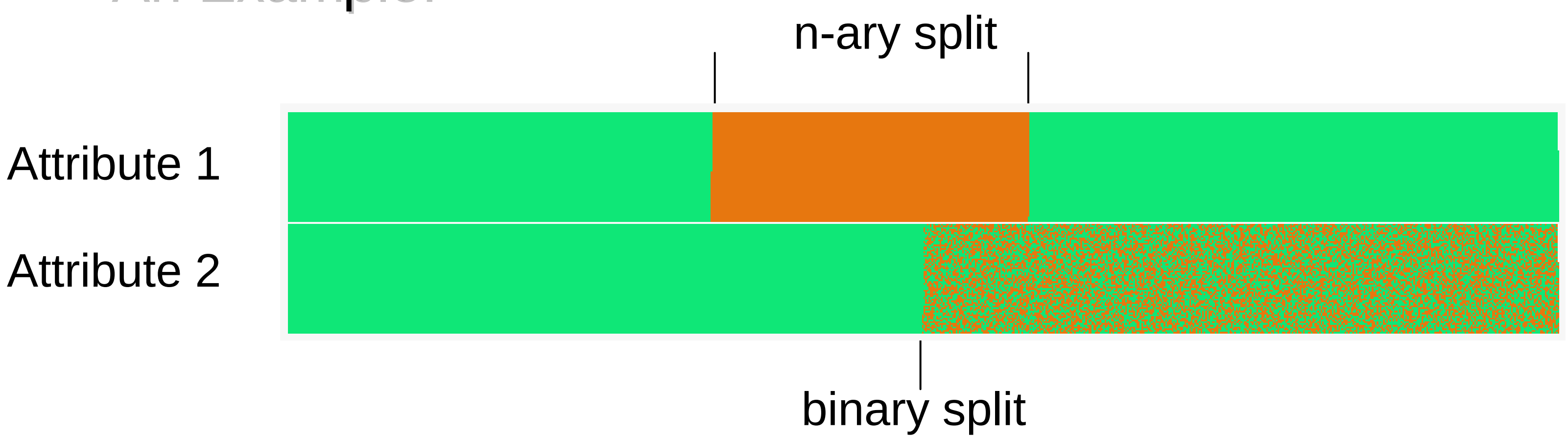
...

- Each attribute is sorted and visualized separately
- Each attribute value is mapped onto a unique pixel
- The color of a pixel is determined by the class label of the object
- The order is reflected by the arrangement of the pixels

Visual Classification

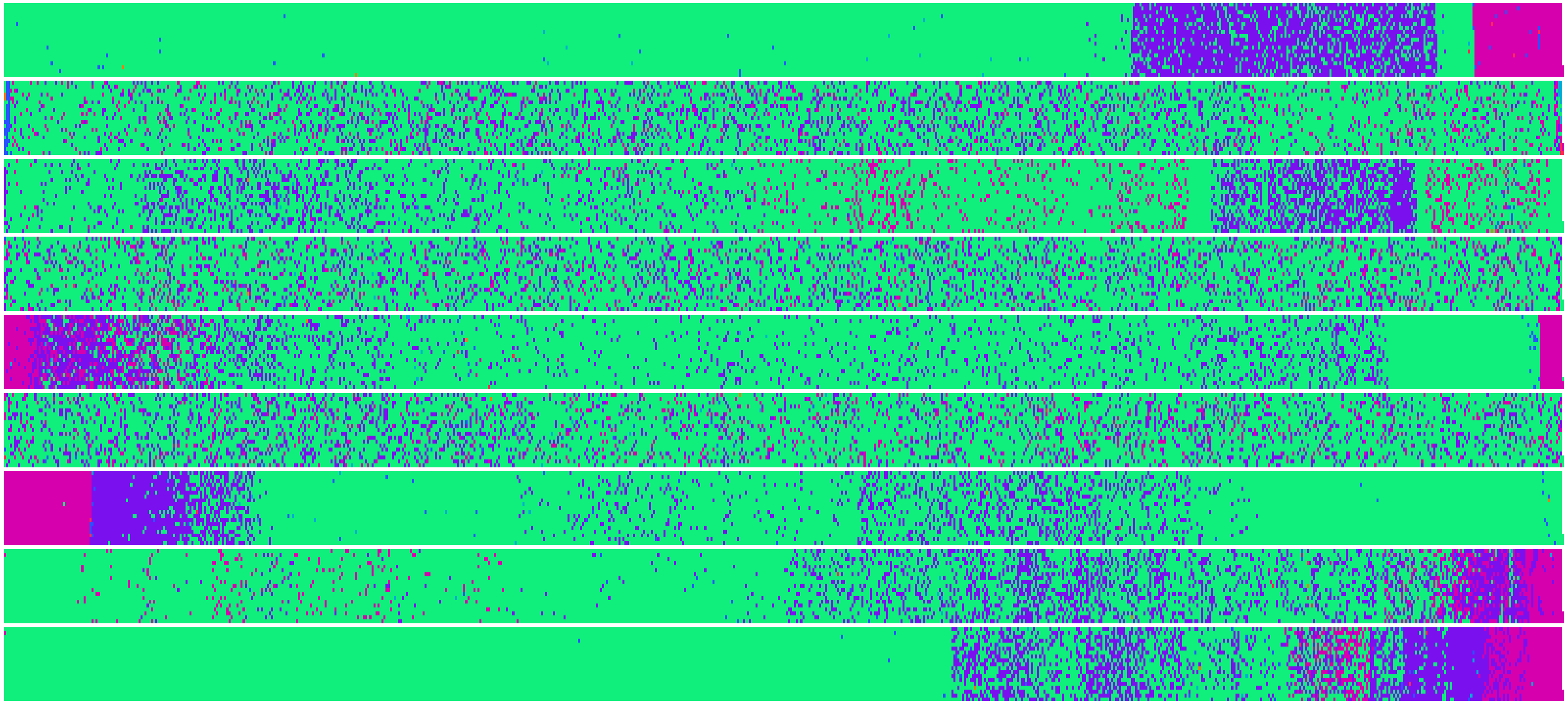


An Example:



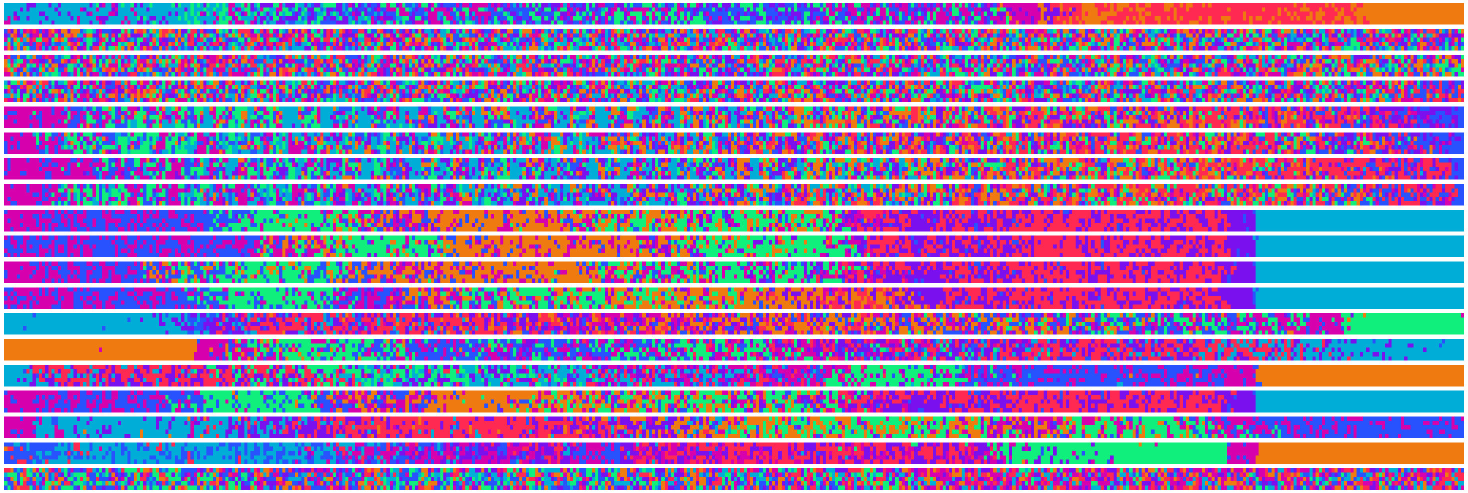
Visual Classification

- Shuttle data set (9 attributes, 43,500 records)



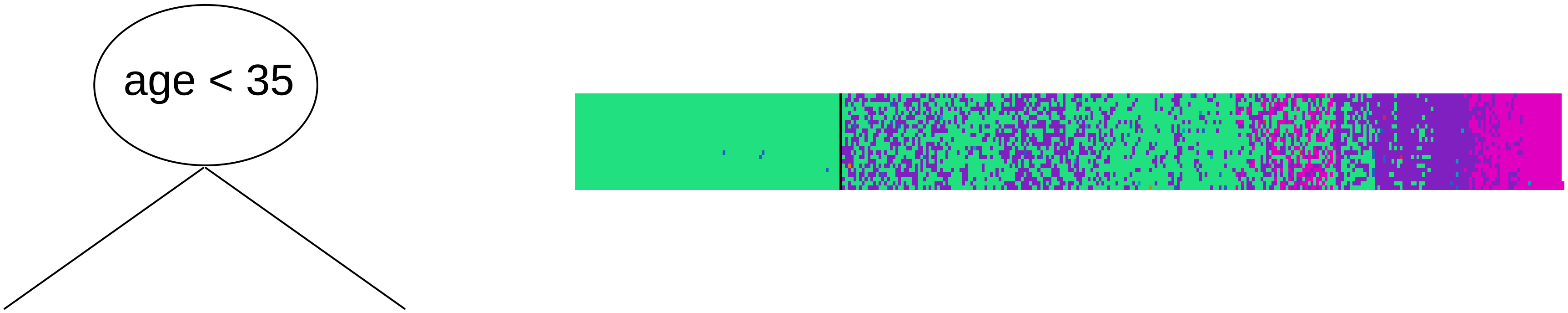
Visual Classification

- Segment data set (19 attributes, 7 classes)



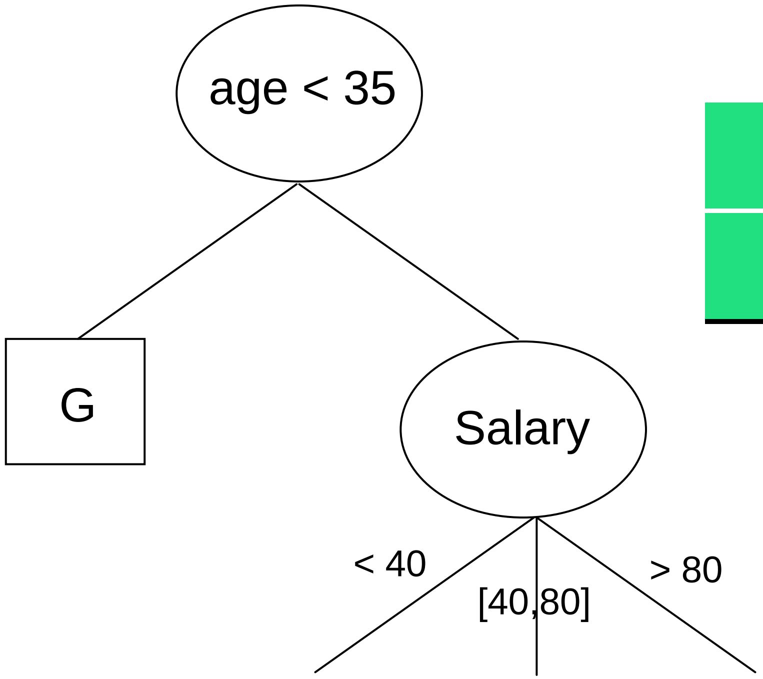
Visual Classification

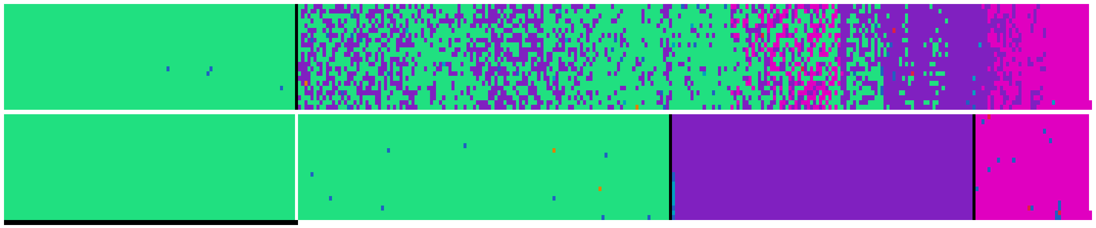
A New Visualization of a Decision Tree



Visual Classification

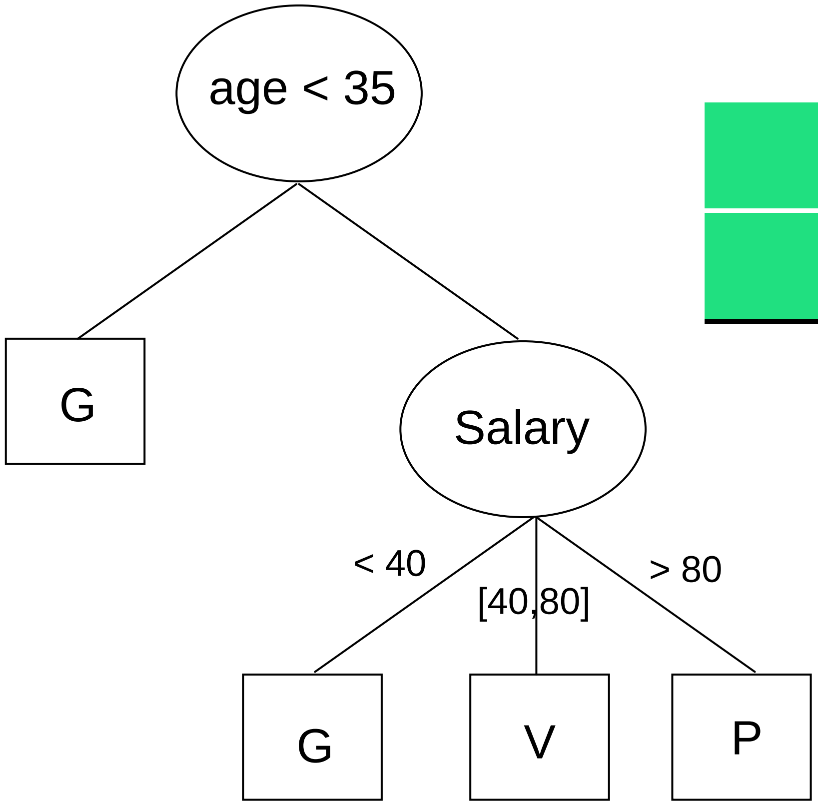
A New Visualization of a Decision Tree

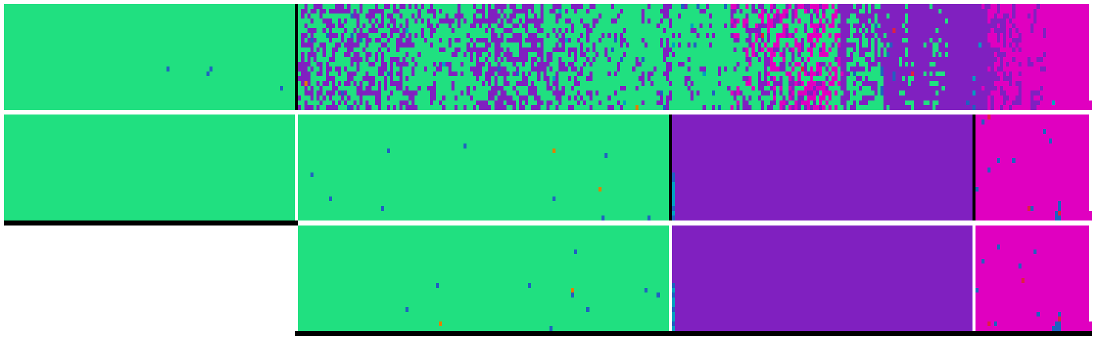




Visual Classification

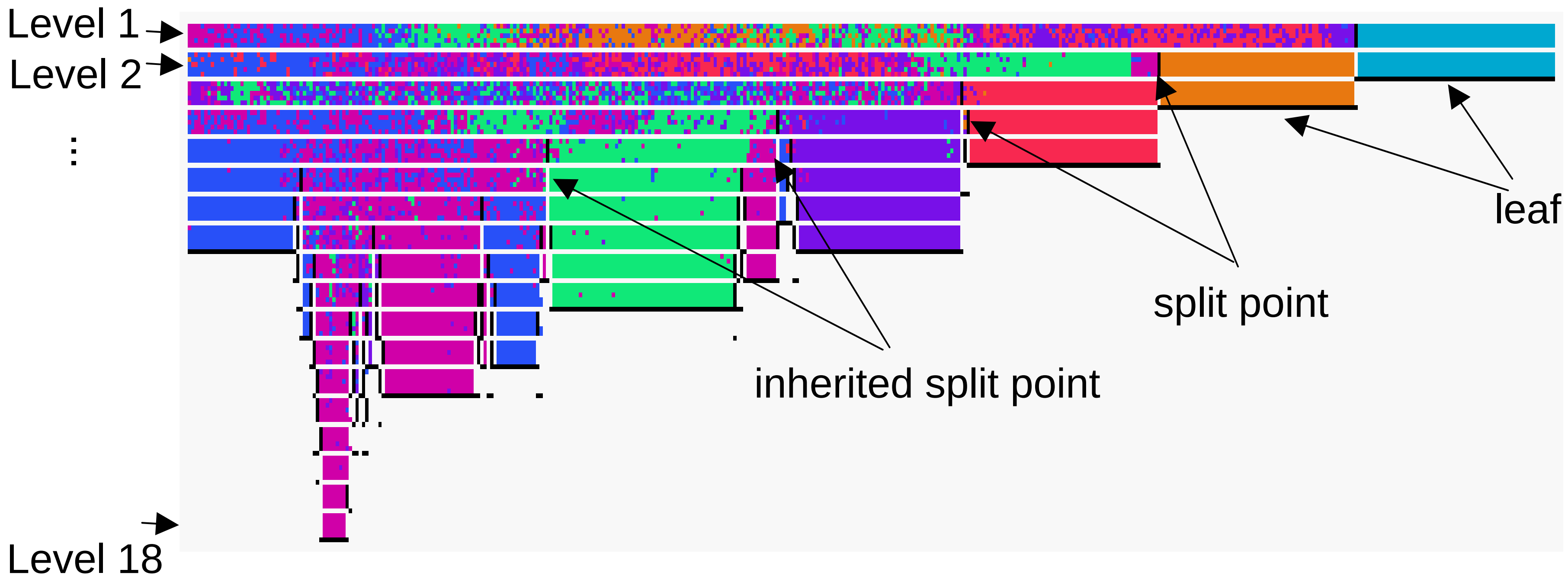
A New Visualization of a Decision Tree





Visual Classification

A Decision Tree for the Segment Data Set



Visual Classification

Different types of algorithmic support for the user:

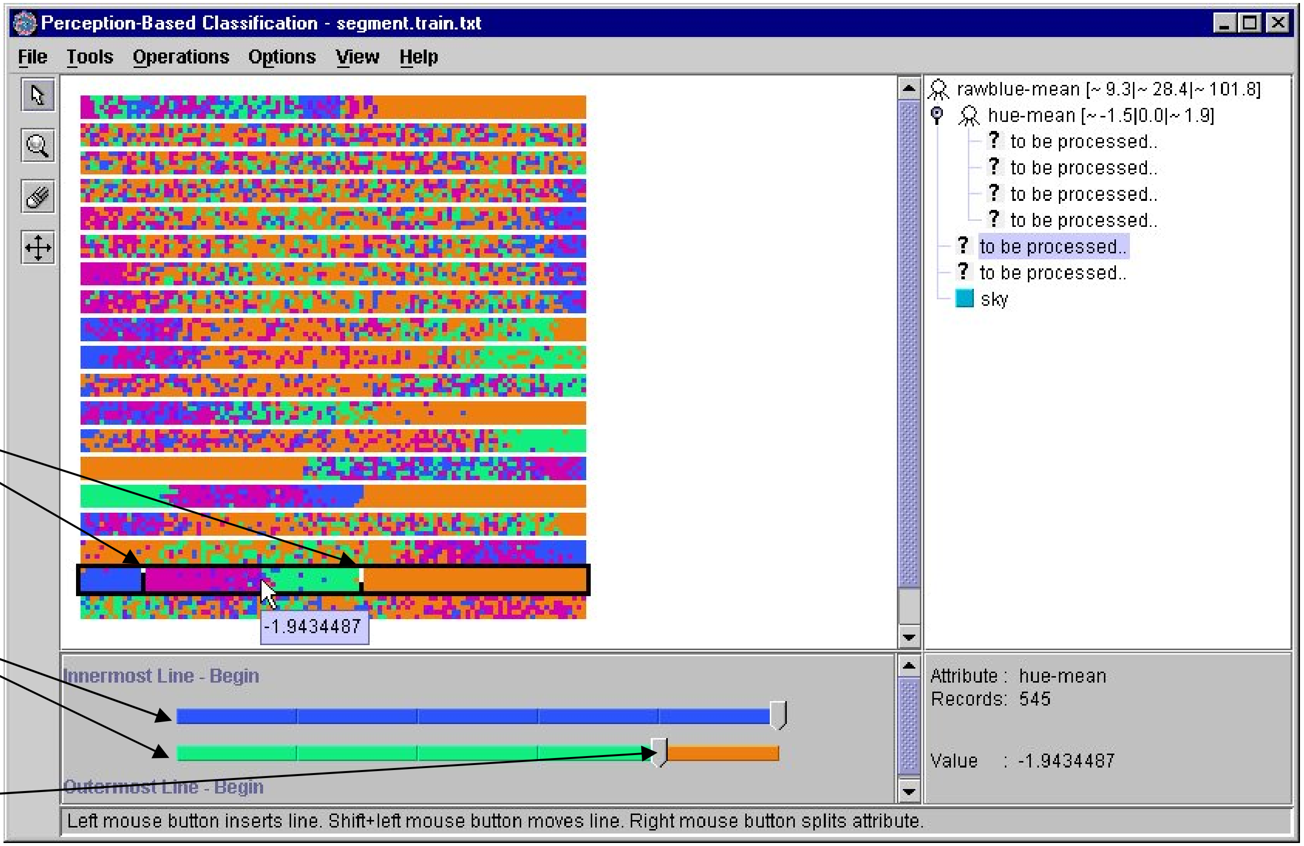
- Propose split
- Look-ahead
- Expand subtree

Visual Classification

animated split lines

magnified split lines

exact split point



Visual Classification

Accuracy:

	Automatic	Automatic-Manual	Manual-Automatic	Manual
Australian	84.9	80.9	86.9	82.7
DNA	93.8	89.2	93.3	89.2
Satimage	86.4	84.1	86.8	83.5
Segment	95.5	95.0	96.3	94.8
Shuttle	99.5	99.6	99.7	99.9

Conclusions

- Tight integration of visualization and data mining algorithms is still a very new area of research

- Data mining algorithms and visualization technique can nicely complement each other.

- PBC leverages decision tree algorithms, allows the user to steer the mining process.

- User involvement during the mining process enables knowledge transfer and capitalizes on human's perception