

**dads 2009 | 2010 spring**

*Theme: Data Mining for Architecture and Urban Planning*

## **Lecture II**

# Fundamental Concepts in Data Mining

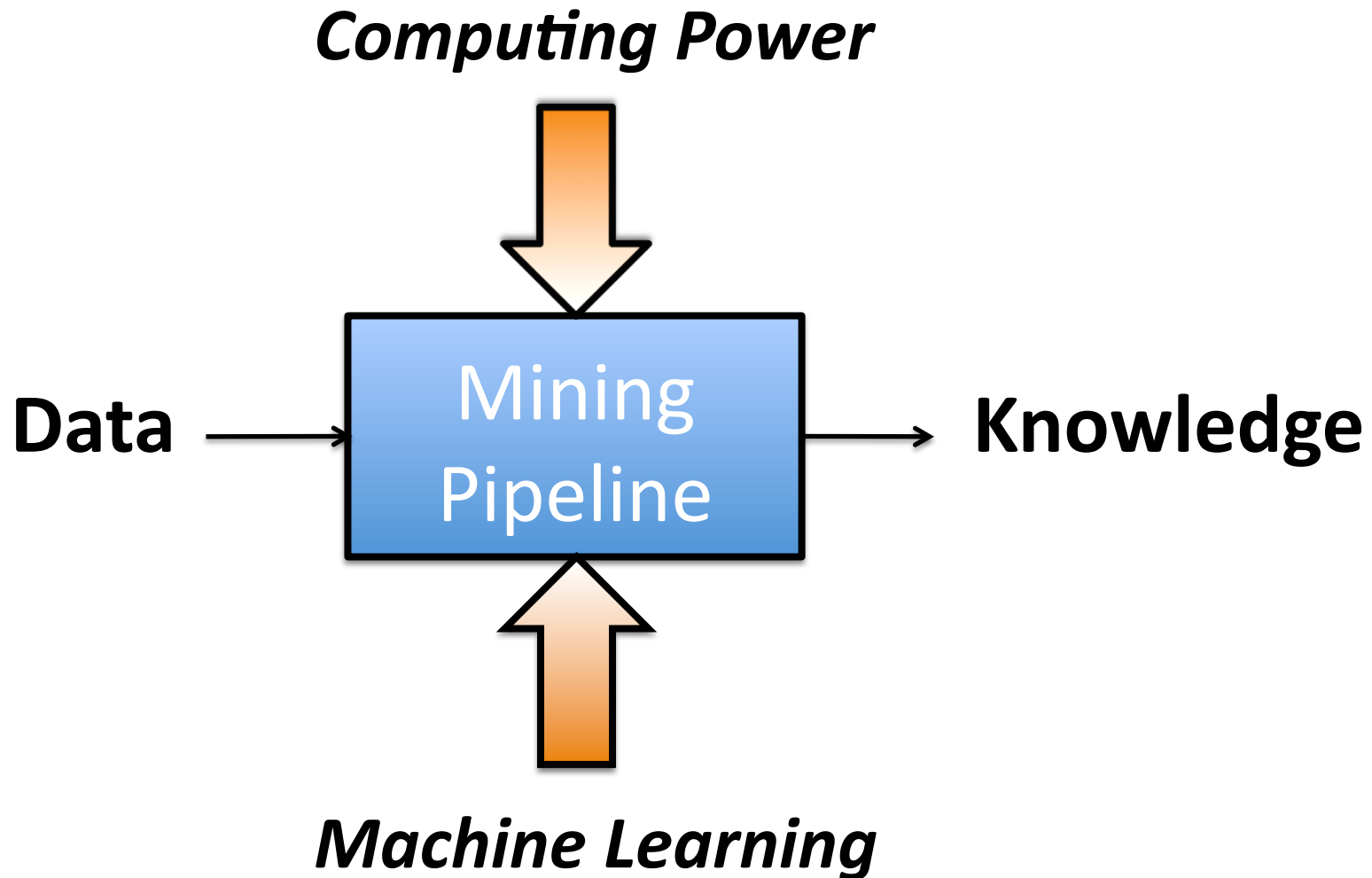
*Ceyhun Burak Akgül, PhD in EE*

*Ahu Sökmenoğlu, M. Arch.*

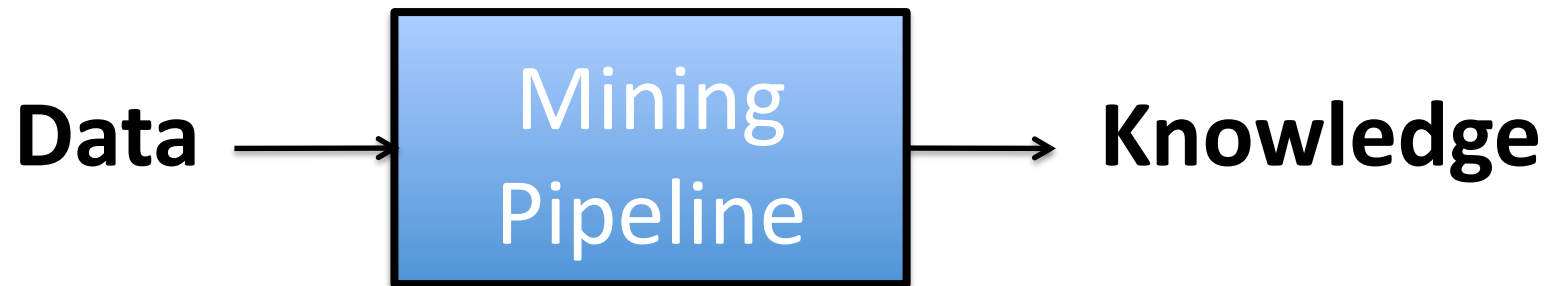
# In this lecture

- Recap: Data mining pipeline
- Inputs – Outputs
- Inputs: Concept Representation
- Outputs: Knowledge Description
- Initiation to RapidMiner

# Data mining pipeline



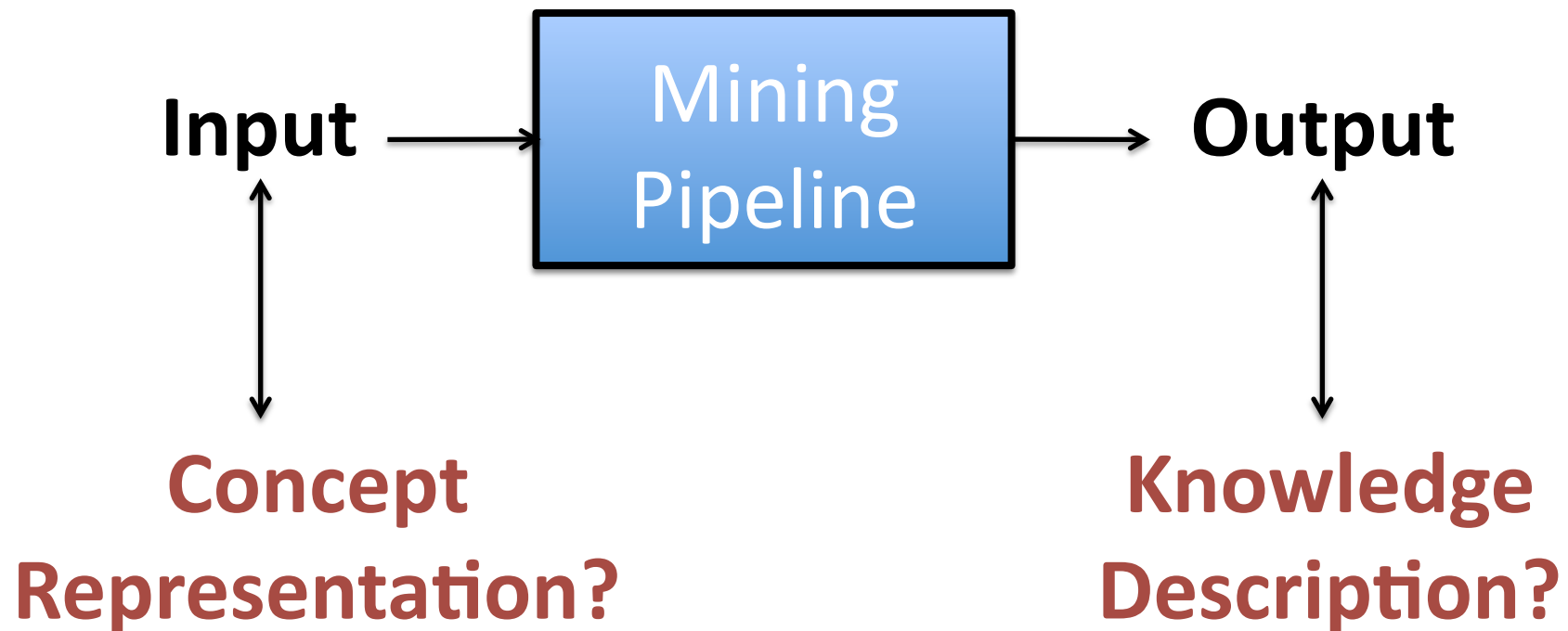
# Data mining pipeline



# Data mining pipeline



# Data mining pipeline



# Inputs: Concept Representation

“What is it in a name? That which we call a rose  
By any other name would smell as sweet”

*Shakespeare*

# Inputs: Concept Representation

- **A concept is an abstraction of a physical thing**  
[Recall Plato's *allegory of cave*]
- Inputs are **instantiations** of a concept



# Inputs: Concept Representation

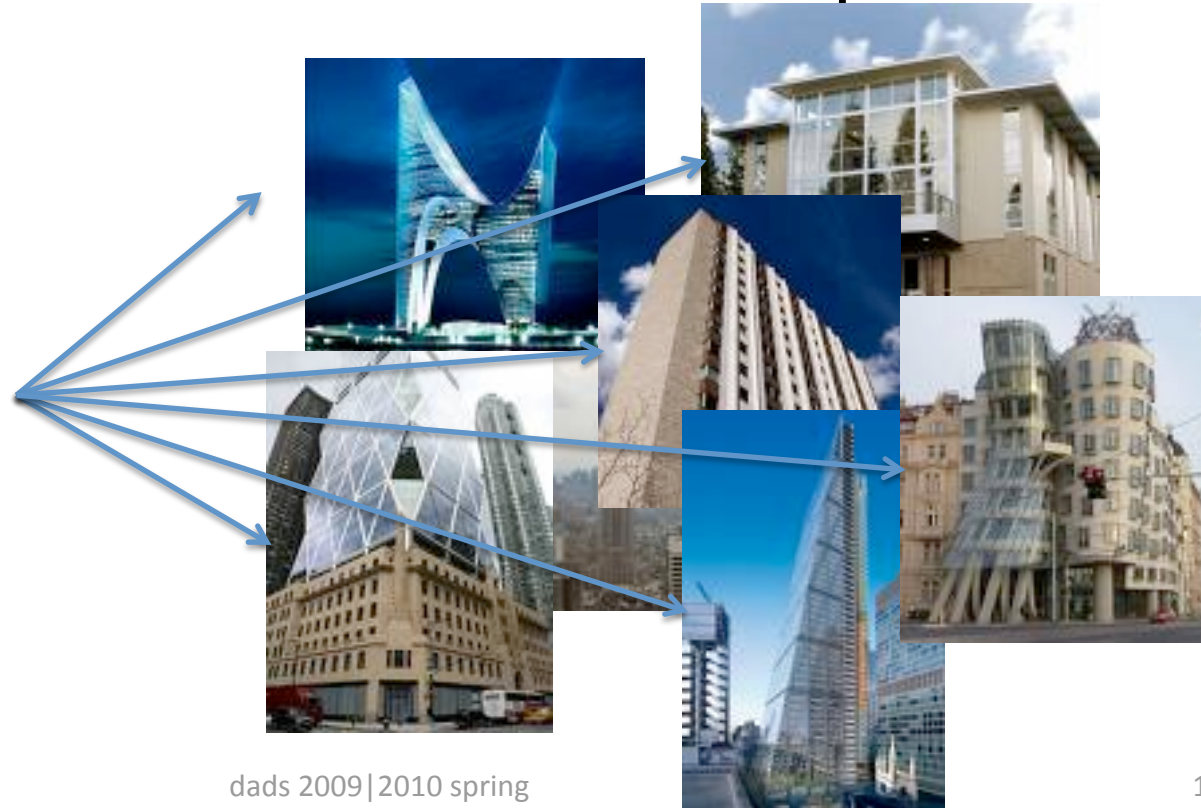
- A concept is an abstraction of a physical thing  
[Recall Plato's *allegory of cave*]
- Inputs are **instantiations** of a concept

## Building

# Inputs: Concept Representation

- A concept is an abstraction of a physical thing  
[Recall Plato's *allegory of cave*]
- Inputs are **instantiations** of a concept

Building



# Inputs: Concept Representation

*How do we represent a concept?*

# Inputs: Concept Representation

*How do we represent a concept?*

Consider the concept “Karnıyarık”

# Inputs: Concept Representation

*How do we represent a concept?*

Consider the concept “Karnıyarık”:

*A traditional Turkish hot dish prepared by stuffing an eggplant with minced meat, finely sliced tomatoes, onion, garlic; flavored with spices ...*

*It tastes good when made with competence...*

*It's a summer dish...*

...

# Inputs: Concept Representation

## *How do we represent a concept?*

Consider the concept “Karnıyarık”:

*A traditional Turkish hot dish prepared by stuffing an eggplant with minced meat, finely sliced tomatoes, onion, garlic; flavored with spices...*

*It tastes good when made with competence...*

*It's a summer dish...*

...

**INFORMATION**

# Inputs: Concept Representation

*How do we represent a concept?*

In data mining:

- We need operational representations
- We need a way to **encode the available information**
- **Attributes** are a means to encode the information
- A concept is thus represented by a set of attributes
- Attributes may not necessarily cover the whole semantic field of the concept\*

# Inputs: Concept Representation

*How do we represent a concept?*

Let's represent “Karnıyarık”:



# Inputs: Concept Representation

*How do we represent a concept?*

Let's represent "Karniyarik":

- Meat
- Eggplant
- Tomato
- Onion
- Garlic
- Spice
- Cooking time
- Taste
- ...



## Attributes

**An attribute is an atomic property of a concept**

# Inputs: Concept Representation

*Input → Concept → Instances*

An example is a particular instance of a concept

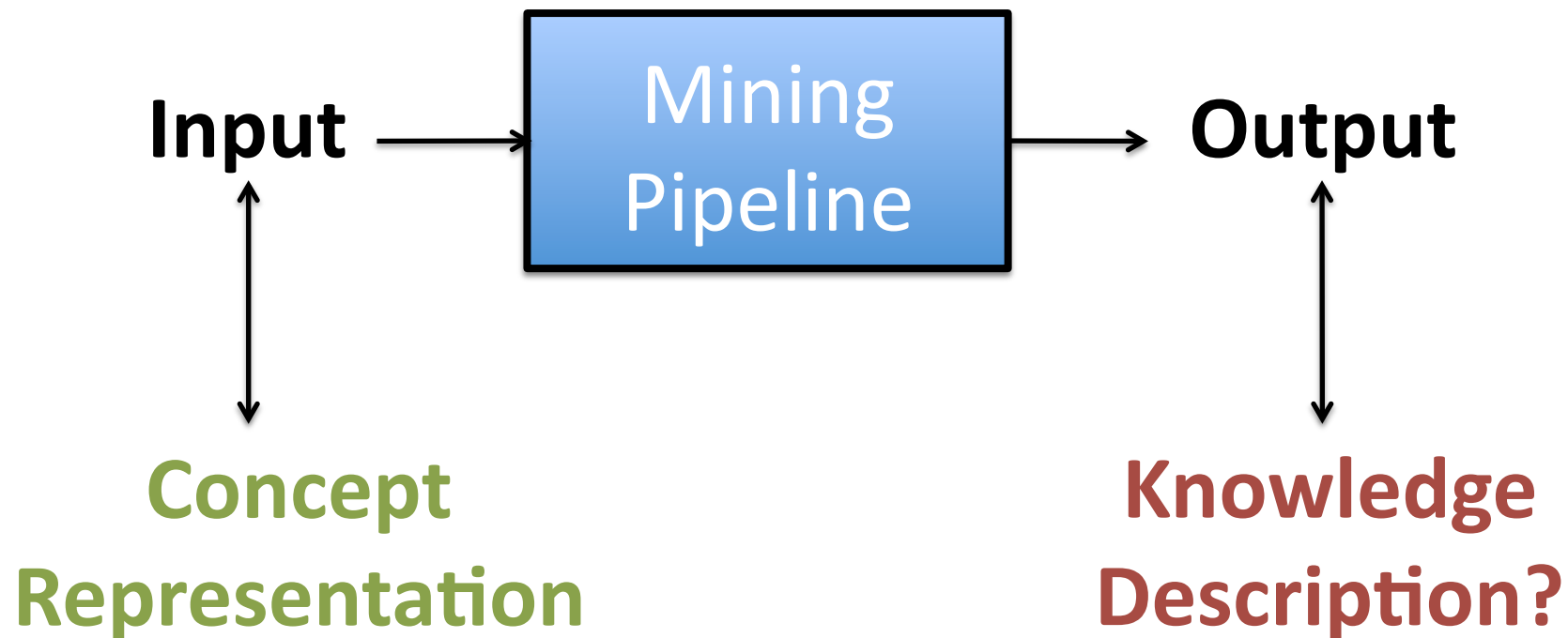
The attributes of an instance are specified

- Meat (grams) = 250 *–numeric*
- Eggplant (piece) = 1 *–numeric*
- Tomato (piece) = 3 *–numeric*
- Onion (piece) = 2 *–numeric*
- Garlic (piece) = 1 *–numeric*
- Spice (spoon) = 1 *–numeric*
- Cooking time (minute) = 30 *–numeric*
- Taste (good/bad/mediocre) = good *–nominal (categorical)*

# Inputs: Concept Representation

- (1) A concept is an abstraction of a physical thing
- (2) Information about a concept is encoded by a set of attributes
- (3) An attribute is an atomic property of a concept
- (4) An instance is a particular realization of a concept
- (5) Input to the data mining pipeline is a set of instances whose attributes are specified

# Data mining pipeline – *again*



# Outputs: Knowledge Description

*How do we describe knowledge?*

# Outputs: Knowledge Description

*How do we describe knowledge?*

- Prediction rules
  - Classification
  - Numeric prediction (Regression)
- Association rules
- Clustering rules

# Outputs: Knowledge Description

*How do we describe knowledge?*

## Prediction rules

- Classification

*Given the amount of each ingredient, cooking time, etc. find a rule predicting whether the “karnıyarık” will taste good, bad, or mediocre.*

- Numeric prediction (Regression)

# Outputs: Knowledge Description

*How do we describe knowledge?*

## Association rules

Any regularity between two or more attributes can be expressed as an association rule

**Ex1.** If meat is X grams, then there should be Y pieces of eggplants such that  $Y = aX$

**Ex2.** If tastes bad, then there were too much onion ( $>>2$ pieces) and cooking time was not enough ( $<10$  mins)

...



# Outputs: Knowledge Description

*How do we describe knowledge?*

**Clustering rules**

# Outputs: Knowledge Description

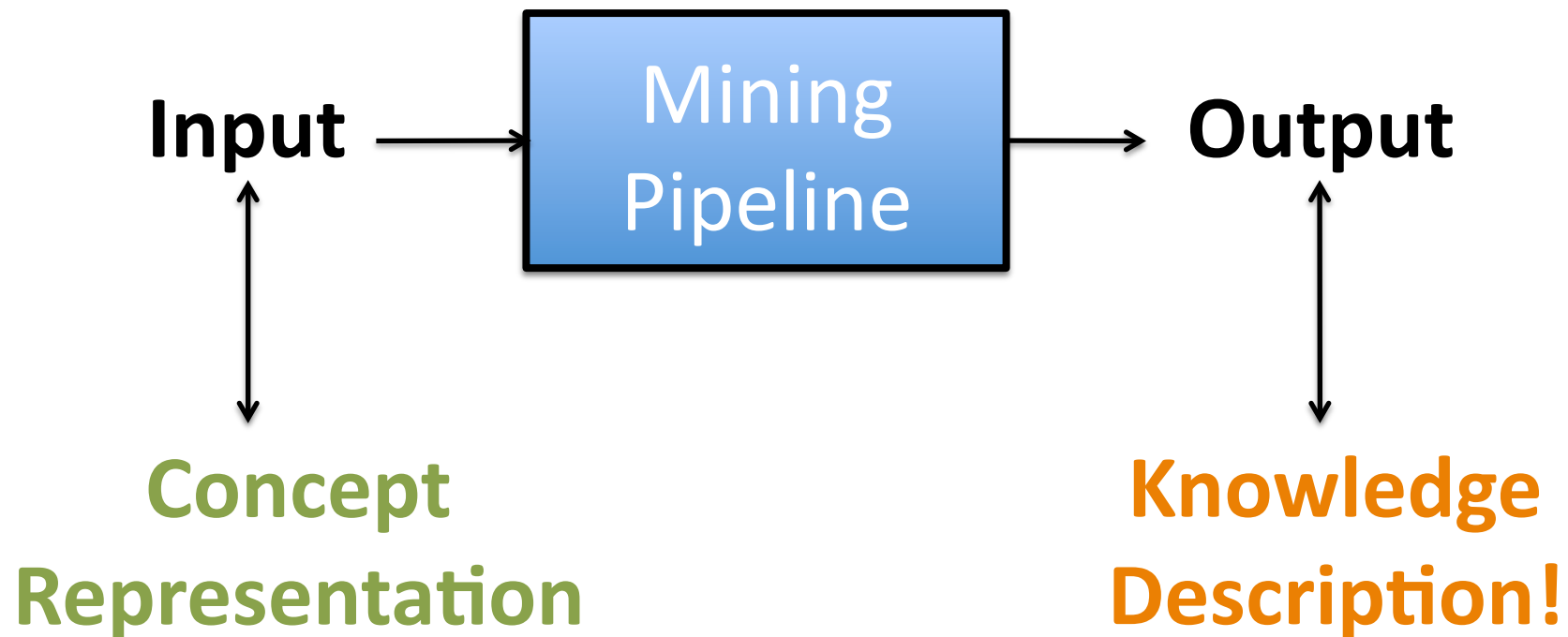
*How do we describe knowledge?*

**Clustering rules**

“Karnıyarık” example is not very instructive!

*See whiteboard* 😊

# Data mining pipeline – *again*



# A Remark

## Where/How do we get data?

Data determines the kind of knowledge  
descriptions you can extract

If you have a specific knowledge description you  
want to obtain, you should collect relevant data  
to this end



# Where are we?

week	date	studio
1	9-Feb	-
2	16-Feb	<b>Introduction: Data Mining in General</b>
3	23-Feb	<b>Concepts in Data Mining</b>
4	2-Mar	Data Mining Applications in Context Introduction to Semester Project
5	9-Mar	Statistics Primer
6	16-Mar	A Broad Picture of Data Mining Tools Jury Meeting; Semester Project's first concepts & ideas
7	23-Mar	Regression and Classification
8	30-Mar	Clustering, Exploratory Data Analysis, and Visualization Semester Project's review
9	6-Apr	Semester Project's review
10	13-Apr	Semester Project's review
11	20-Apr	Jury Meeting; Presentations
12	27-Apr	Semester Project's review
13	4-May	Semester Project's review
14	11-May	Jury Meeting; Final Presentations

# Assignments for next week

## Think of the “City” as a concept:

- Designate a set of attributes related to the city
- Instantiate the “city” concept with several examples
- Specify the attributes of your “city” examples

**What kind of knowledge descriptions can you extract with your chosen set of attributes?**

**Do the reverse\***

# Assignment for two weeks from now

**Download and install RapidMiner**

<http://rapid-i.com/content/view/181/190/>

**Try to figure out what it is about 😊**