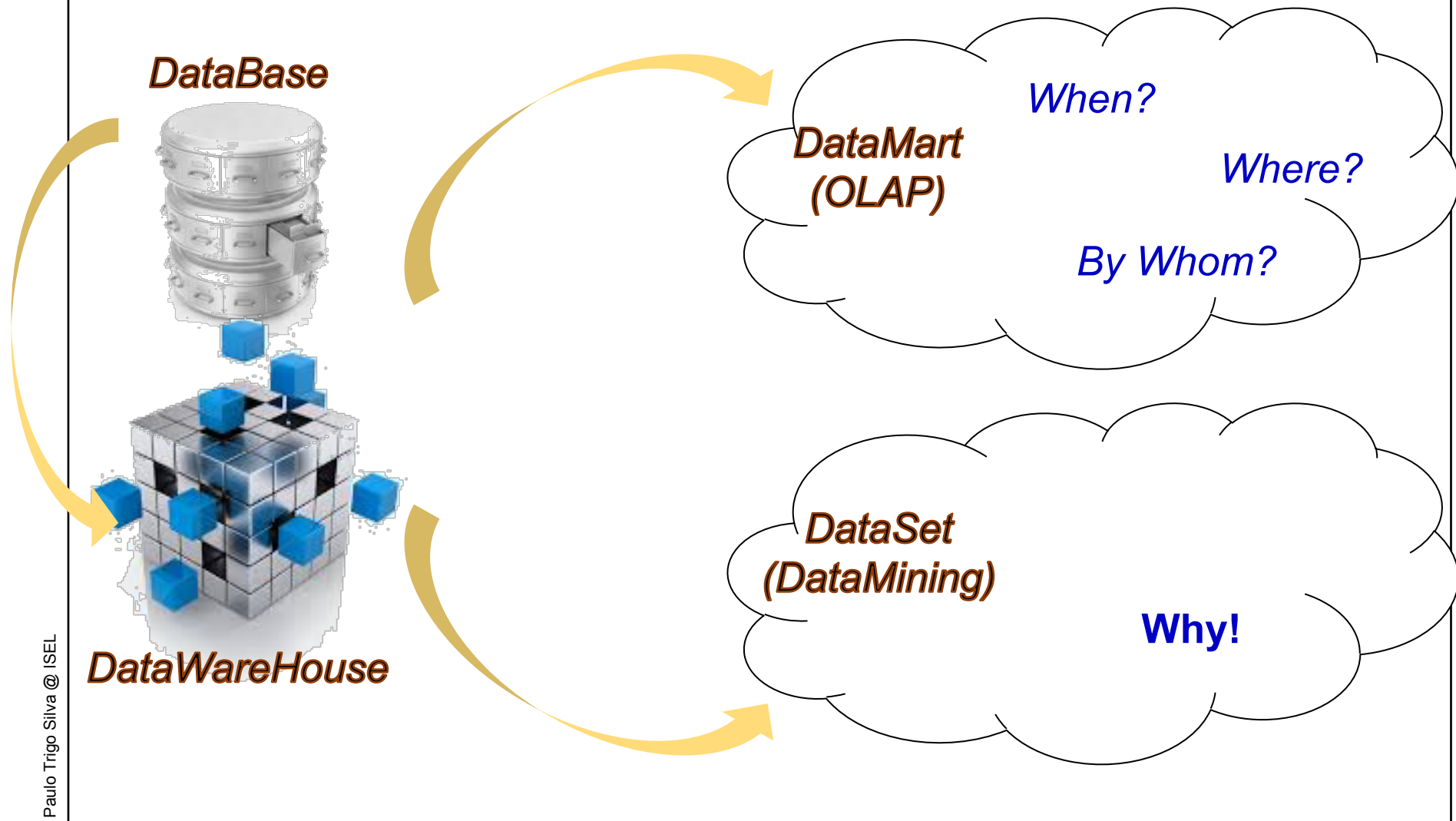


Knowledge Extraction

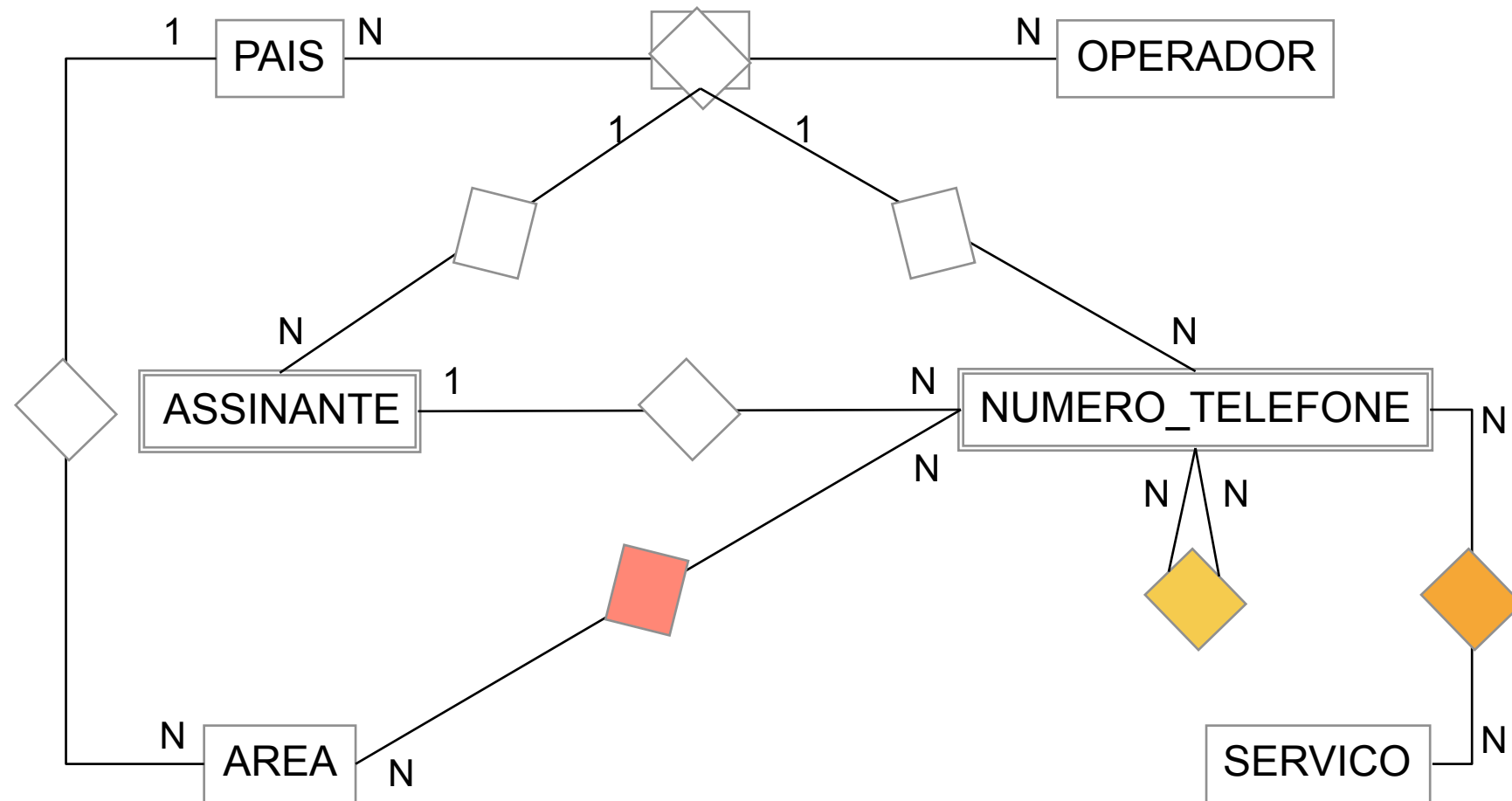
The data, the formats and the questions placed on them



... the formats – different for different goals

<i>DataBase</i>	<i>DataWareHouse</i>	<i>DataMining</i>
SEVERAL-TABLES	SEVERAL-CUBES	ONE-TABLE
<i>Relacional-Schema</i>	<i>Data-Mart</i>	<i>DataSet</i>
<u>query and manipulation of data via SQL</u>	<u>exploitation of aggregated data via MDX</u>	<u>extraction of “tendencies” via algorithms</u>

Relational-Schema (via Entity-Relationship model)



represents the concept of "localization"



represents the concept of "billing"

represents the concept of "profile"

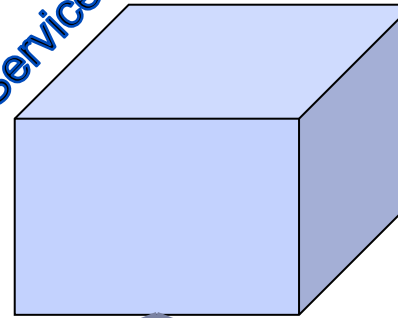
DataMart

hipercube

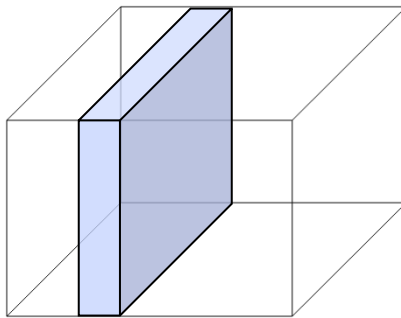
Operator

Service

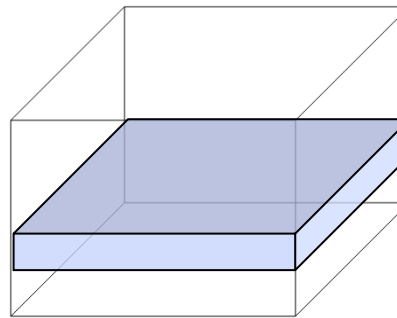
Country



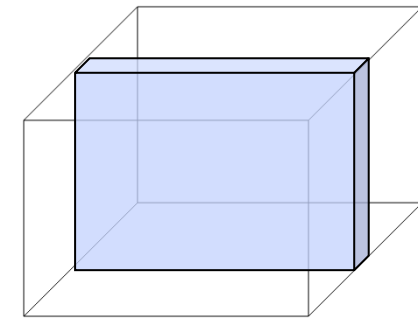
selection by the
value of a dimension
(slice)



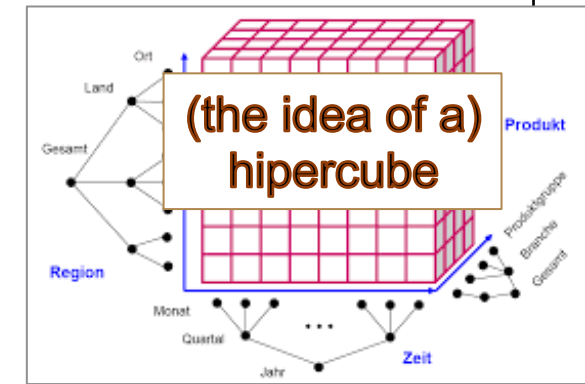
all countries and
services of an
operator



all services and
operatos of a
country



all countries and
operators with a
serviço



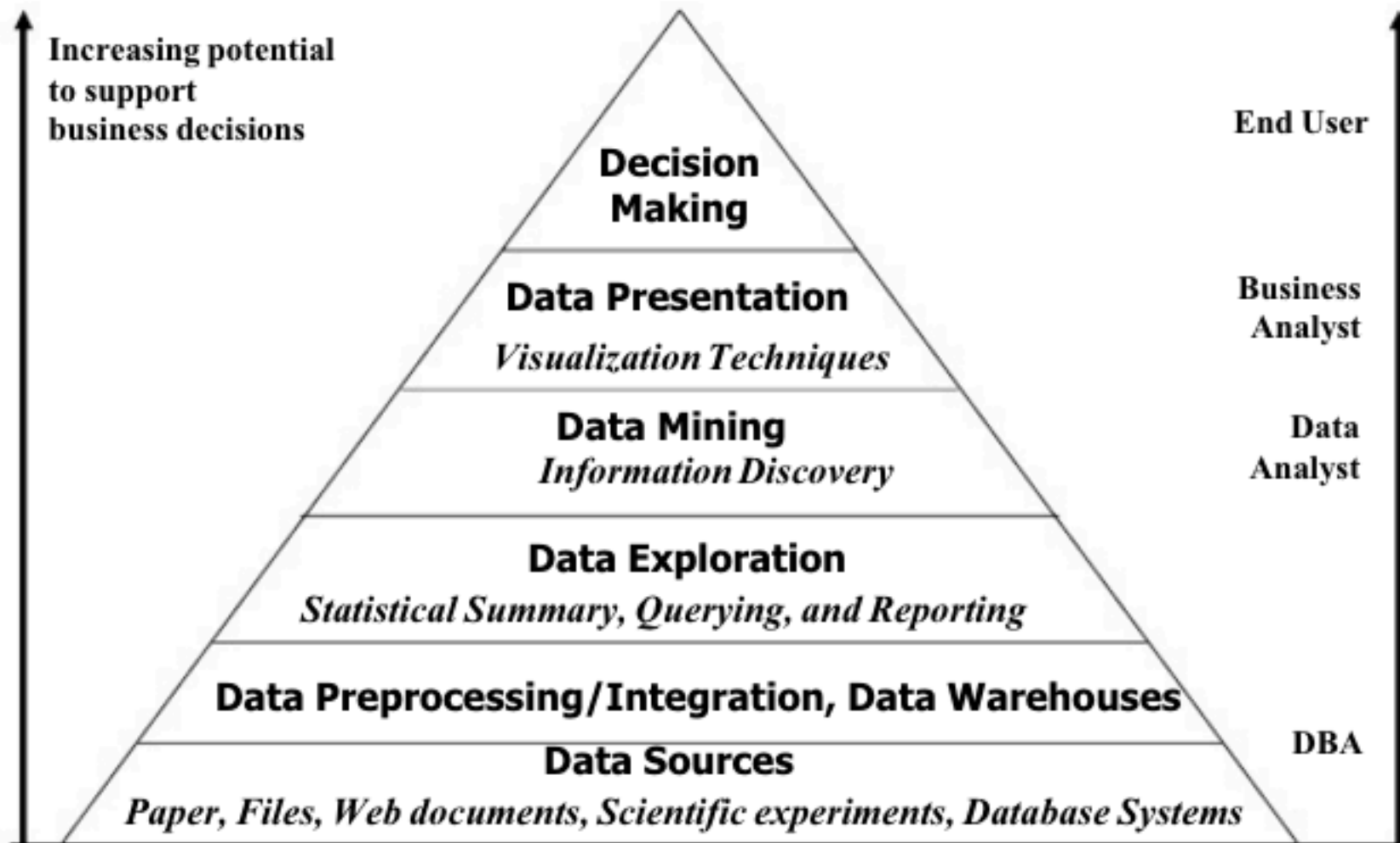
DataSet

The data is represented as:

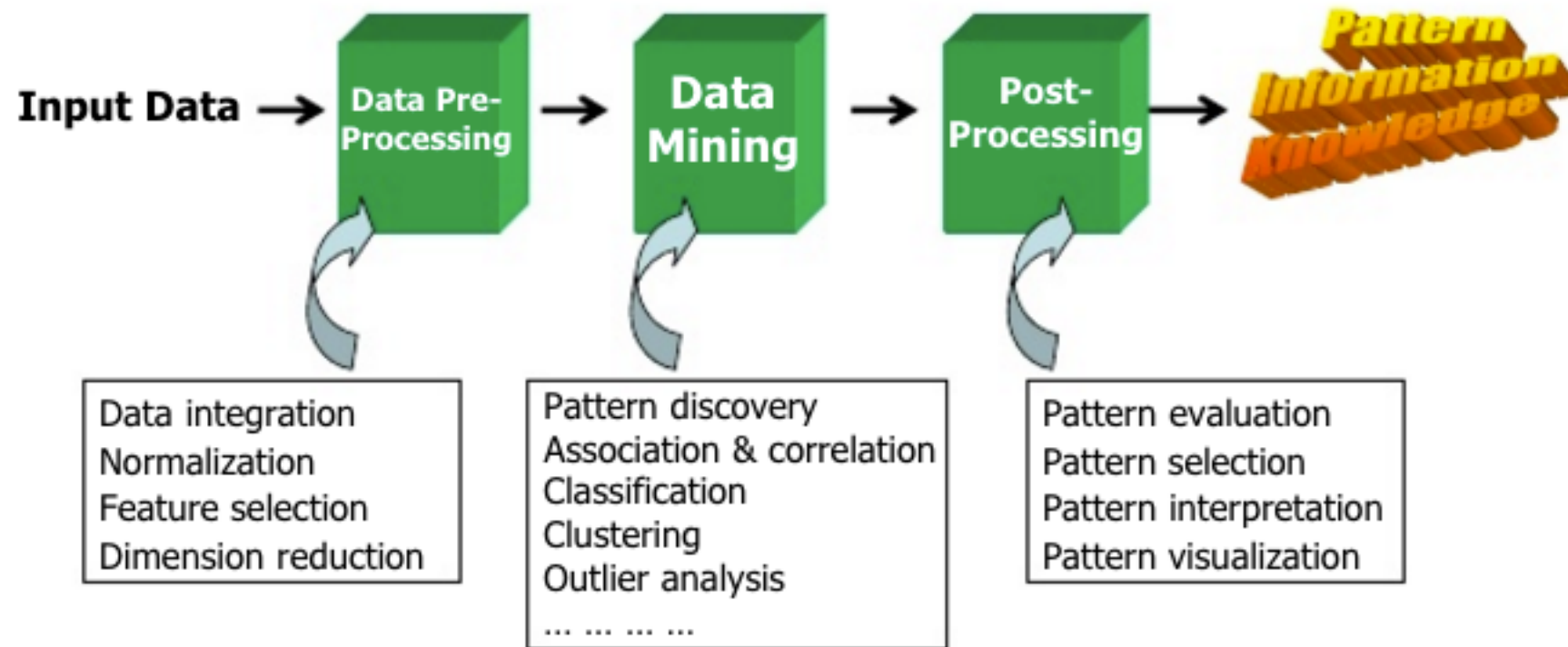
- a “single table”, i.e., a unique and “huge” vector with D -dimensions,
- each “column” represents a characteristic (“*feature*”) of what is intended to be analyzed,
- each “column” (“*feature*”) has values that vary according to their type.

The screenshot shows a web browser window with the address bar displaying 'syriaw_1_shoes.com/bsol'. The main content area contains a table with 10 columns and 15 rows of data. The columns are labeled as follows: 'Product Name', 'Price', 'Quantity', 'Status', 'Date', 'Time', 'Location', 'Category', 'Sub-category', and 'Brand'. The data is organized into a grid with alternating light and dark rows. The first row of data is: 'Product Name: ...', 'Price: ...', 'Quantity: ...', 'Status: ...', 'Date: ...', 'Time: ...', 'Location: ...', 'Category: ...', 'Sub-category: ...', 'Brand: ...'. The table continues with similar data for 14 more rows. The browser's address bar and the table's header row are highlighted in blue.

Data-Mining and Business-Intelligence \ integrated view



KDD – Knowledge Discovery in Databases \ the process



this is a view from typical machine learning and statistics communities

Data Mining – Confluence of Multiple Disciplines

