Zagazig University

Faculty of Computers and Informatics

Mining Association Rule

Supervisored By :

Prof : Abd_Elazem Negm

Eng: Nora Zaki

Group Members

- Tawheda Hassan El_Shoura
- Rahma Moaz Mohammed
- Sara Abd_Allah Hafez
- Raghda Reda Salem
- Seham Mohammed El_Bayomy

Data Mining
Concepts & Techniques

Data Mining Concepts

What is data mining?

- Data mining refers to: extracting or ”mining” knowledge from large amounts of data.

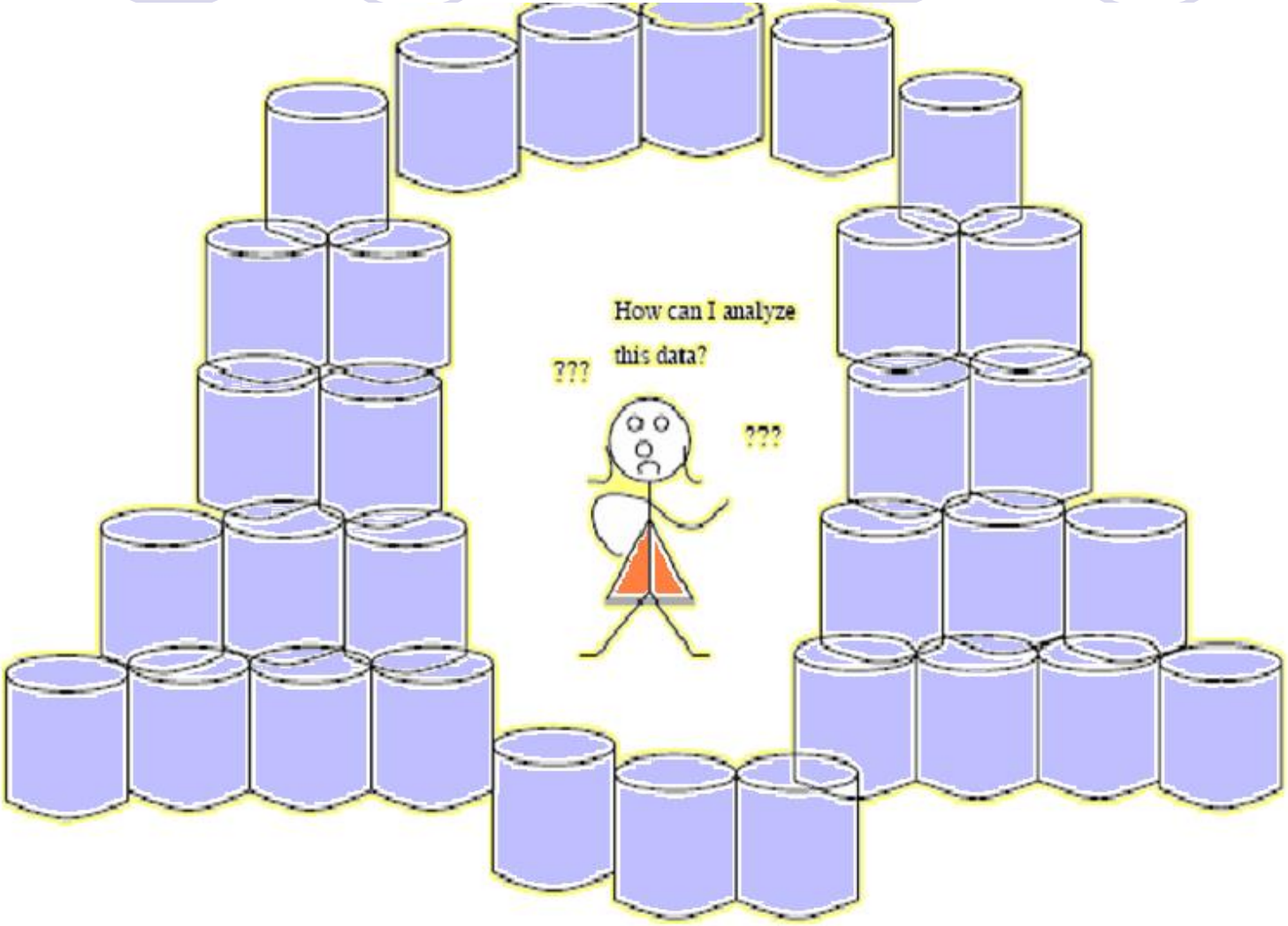
- Data mining is the task of discovering interesting patterns from large amounts of data where the data can be stored in databases, data warehouses, or other information repositories.

Data Mining Concepts(cont.)

Example :

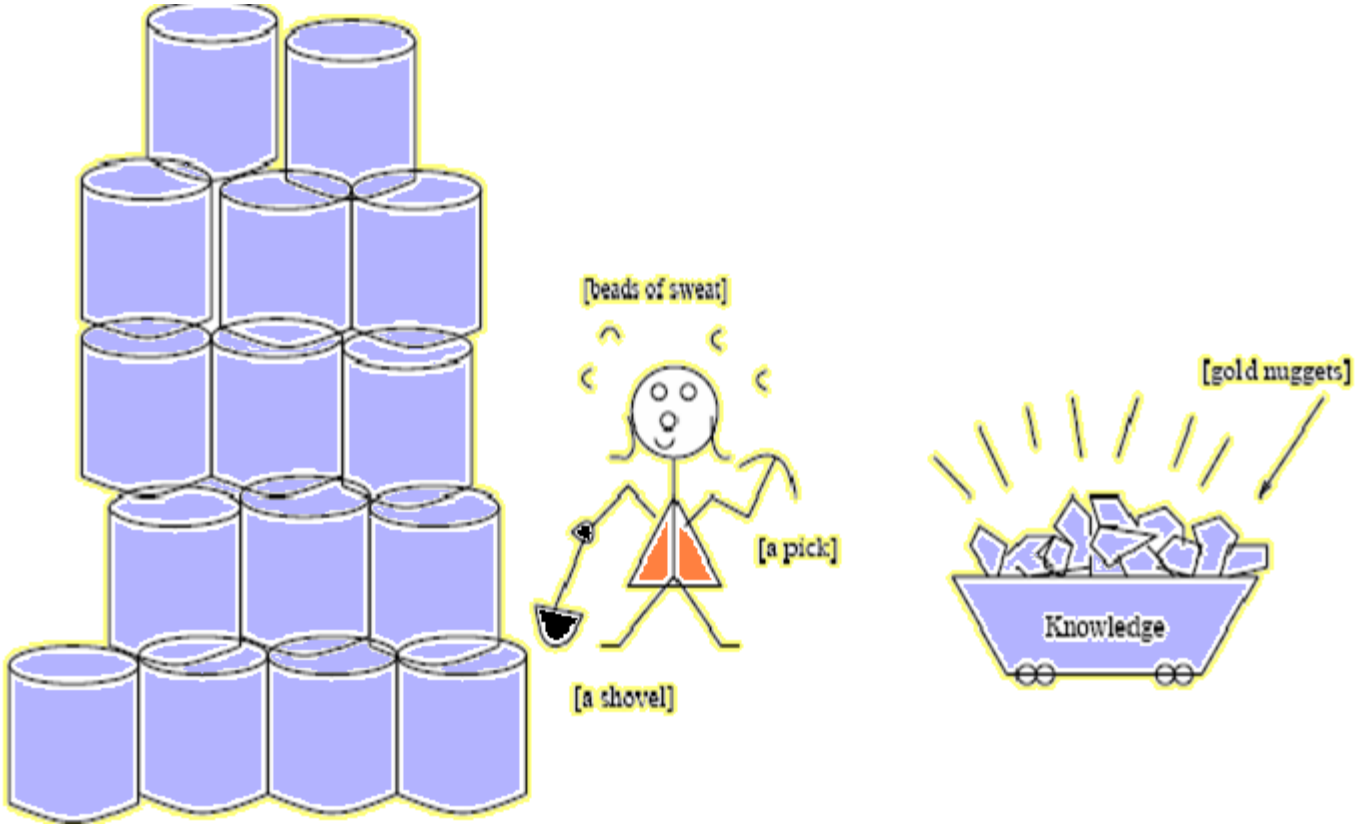
- Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus," data mining" should have been more appropriately named \knowledge mining from data", which is unfortunately somewhat long. Knowledge mining", a shorter term, may not reflect the emphasis on mining from large amounts of data.

Data Miming Concepts



We have rich information ,Buti Information poor.

Data Miming Concepts(cont.)



Data mining - searching for knowledge (interesting patterns) in your data

Data Miming Concepts(cont.)

Data Mining Steps

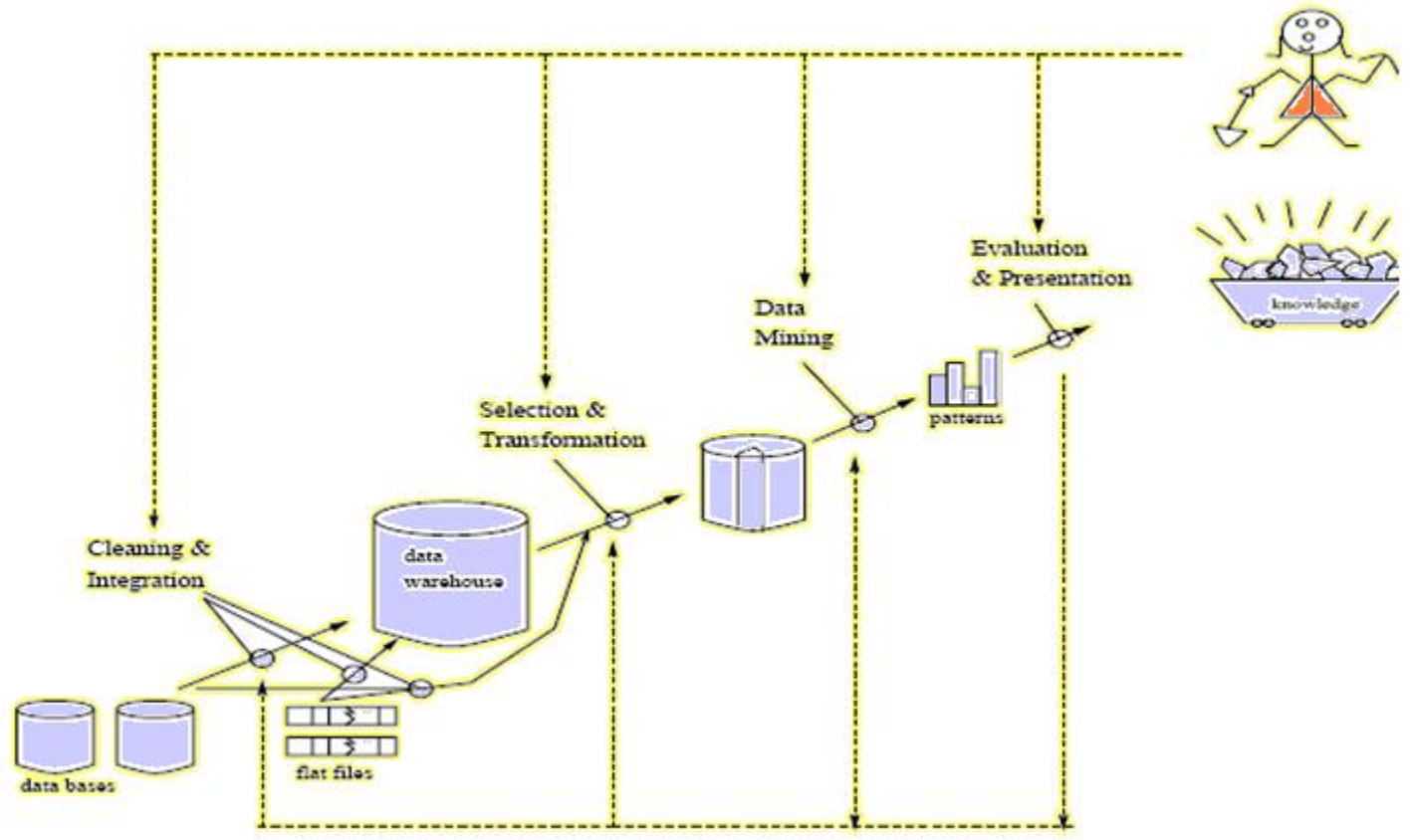
- data cleaning (to remove noise or irrelevant data).
- data integration (where multiple data sources may be combined).
- data selection (where data relevant to the analysis task are retrieved from the database).
- data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance).

Data Miming Concepts(cont.)

Data Mining Steps(cont.)

- data mining (an essential process where intelligent methods are applied in order to extract data patterns).
- pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures.
- knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user).

Data Miming Concepts(cont.)



Data Mining as process of knowledge discovery

Data Mining Concepts(cont.)

- What Kind Of Data?

  - Relational databases
  - Data warehouses
  - Transactional databases
  - Advanced database systems and advanced database applications

Data Mining Concepts(cont.)

- What are data mining functionalities?

- Concept/class description: characterization and discrimination
- Association Rule
- Classification and prediction
- Clustering analysis
- Evolution and deviation analysis

Association Rule

Association Rule

- Find all frequent itemsets
- Generate strong association rules from the frequent item sets
- searches for interesting relationships among items in a given data set.

Association Rules

- Based on the types of values, the association rules can be classified into two categories : Boolean Association Rules and Quantitative association Rules.

- ***Boolean Association Rule***: Keyboard Mouse
  [support = 6%,confidence= 70%]

- ***Quantitative Association Rule***: (Age = 26…30) ⇒ (Cars =1, 2) [Support 3%,confidence =36%]

Minimum Confidence Threshold

- Confidence is defined as the measure of certainty or trustworthiness associated with each discover pattern .

$$If\ A \Longrightarrow B$$

$$Confidence(A \Longrightarrow B) = \frac{\#\_Tuples\_containing\_both\_A\_and\_B}{\#\_Tuples\_containing\_A}$$

Apriori Algorithm

Apriori Algorithm

- Find frequent itemsets using an iterative level wise approach
- The Apriori Algorithm : is an influential algorithm for mining frequent itemsets for boolean association rules.

The Apriori Algorithm: Basics

- Key Concepts :

  - Frequent Itemsets: The sets of item which has minimum support (denoted by Lifor ith-Itemset).

  - Apriori Property: Any subset of frequent itemset must be frequent.

  - Join Operation: To find Lk, a set of candidate k-itemsets is generated by joining Lk-1with itself.

The Apriori Algorithm Steps

- Find the *frequent itemsets*: the sets of items that have minimum support .
- A subset of a frequent itemset must also be a frequent itemset.
  i.e., if {*AB*} isa frequent itemset, both {*A*} and {*B*} should be a frequent itemset .
- Iteratively find frequent itemsets with cardinality from 1 to *k* *(k*-itemset*)* .
- Use the frequent itemsets to generate association rules.

The Apriori Algorithm : Pseudo code

- Join Step: Ckis generated by joining Lk-1with itself
- Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset
- Pseudo-code:

Ck: Candidate itemset of size k

Lk: frequent itemset of size k

L1= {frequent items};

for (*k*= 1; *Lk*!=∅; *k*++) **do begin**

Ck+1= candidates generated from *Lk*;

for each transaction *t*in database do

increment the count of all candidates in *Ck+1*that are contained in *t*

Lk+1= candidates in *Ck+1*with min_support

end

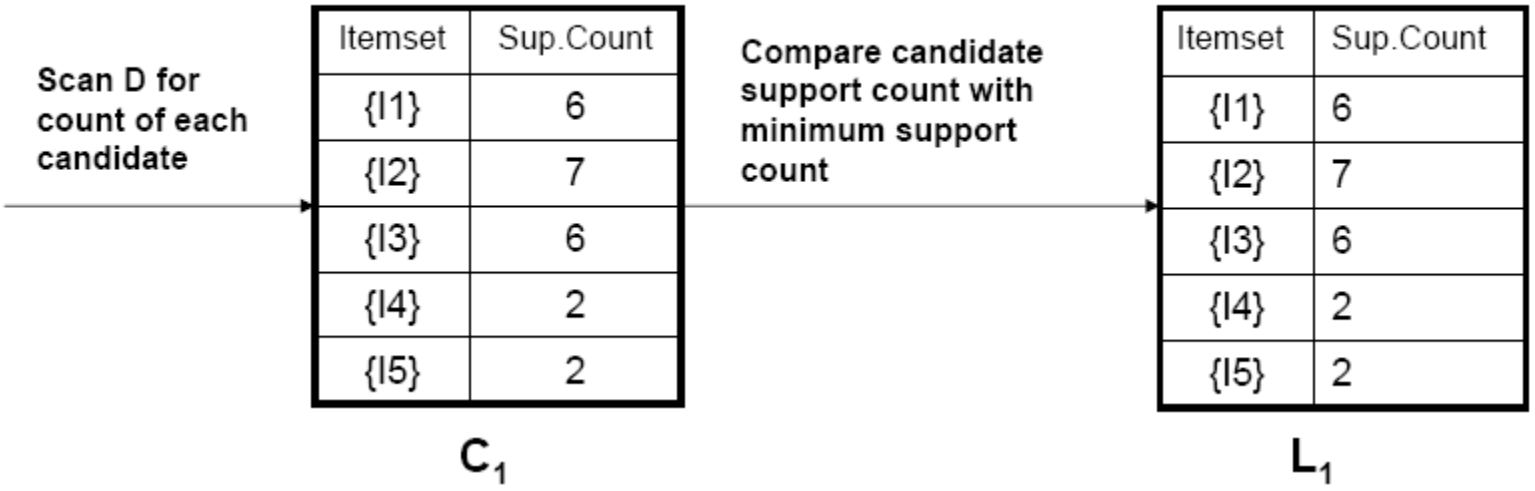
return *kLk*;

The Apriori Algorithm: Example

TID	List of Items
T100	I1, I2, I5
T100	I2, I4
T100	I2, I3
T100	I1, I2, I4
T100	I1, I3
T100	I2, I3
T100	I1, I3
T100	I1, I2 ,I3, I5
T100	I1, I2, I3

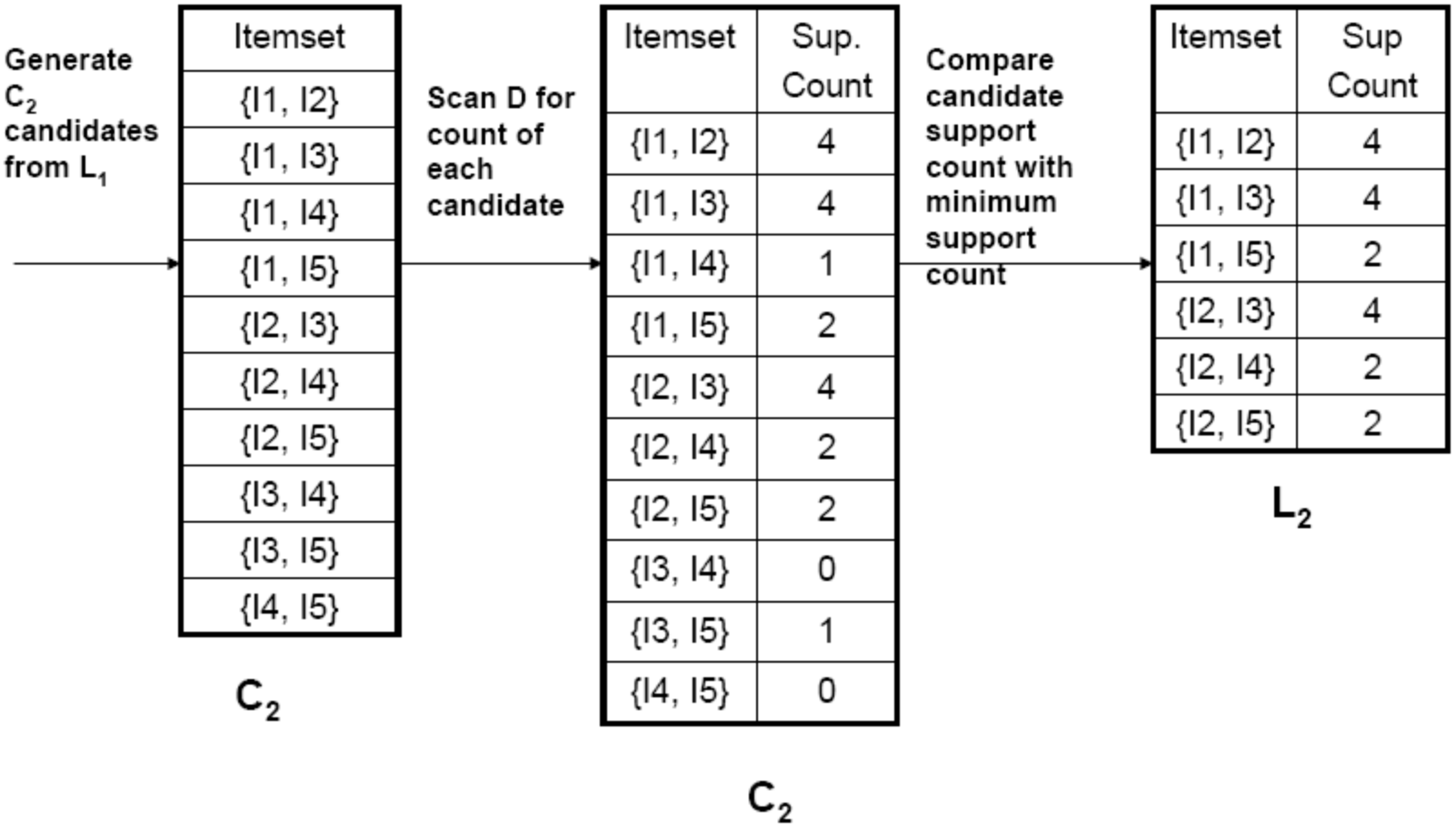
- Consider a database, D , consisting of 9 transactions.
- Suppose min. support count required is 2 (i.e. min_sup = 2/9 = 22 %)
- Let minimum confidence required is 70%.
- We have to first find out the frequent itemset using Apriori algorithm.
- Then, Association rules will be generated using min. support & min. confidence.

Step 1: Generating 1-itemset Frequent Pattern



- The set of frequent 1-itemsets, L_1 , consists of the candidate 1-itemsets satisfying minimum support.
- In the first iteration of the algorithm, each item is a member of the set of candidate.

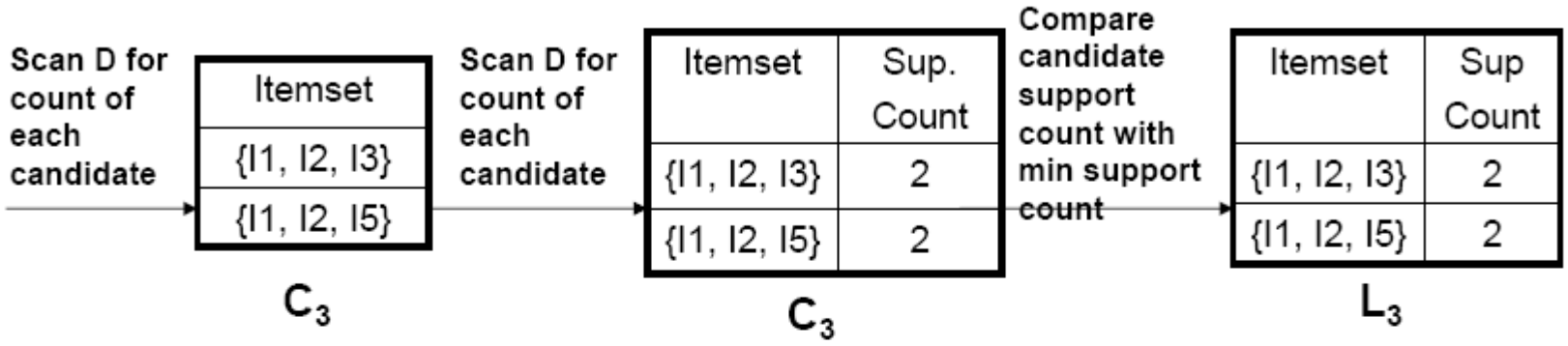
Step 2: Generating 2-itemset Frequent Pattern



Step 2: Generating 2-itemset Frequent Pattern

- To discover the set of frequent 2-itemsets, L2, the algorithm uses L1 *Join* L1to generate a candidate set of 2-itemsets, C2.
- Next, the transactions in D are scanned and the support count for each candidate itemset in C2is accumulated (as shown in the middle table).
- The set of frequent 2-itemsets, L2, is then determined, consisting of those candidate 2-itemsets in C2having minimum support.
- Note:We haven't used Apriori Property yet.

Step 3: Generating 3-itemset Frequent Pattern



- The generation of the set of candidate 3-itemsets, C_3 , involves use of the Apriori Property.
- In order to find C_3, we compute L_2 *Join* L_2.
- C_3 = L2 *Join* L2 = {{I1, I2, I3}, {I1, I2, I5}, {I1, I3, I5}, {I2, I3, I4}, {I2, I3, I5}, {I2, I4, I5}}.
- Now, Join step is complete and Prune step will be used to reduce the size of C_3. Prune step helps to avoid heavy computation due to large C_k.

Step 3: Generating 3-itemset Frequent Pattern

- Based on the Apriori property that all subsets of a frequent itemset must also be frequent, we can determine that four latter candidates cannot possibly be frequent. How ?

- For example , lets take {I1, I2, I3}.The 2-item subsets of it are {I1, I2}, {I1, I3} & {I2, I3}. Since all 2-item subsets of {I1, I2, I3} are members of L2, We will keep {I1, I2, I3} in C3.

- Lets take another example of {I2, I3, I5}which shows how the pruning is performed. The 2-item subsets are {I2, I3}, {I2, I5} & {I3,I5}.

Step 3: Generating 3-itemset Frequent Pattern (cont.)

- BUT, {I3, I5} is not a member of L2 and hence it is not frequent violating Apriori Property. Thus We will have to remove {I2, I3, I5} from C3.
- Therefore, C3= {{I1, I2, I3}, {I1, I2, I5}} after checking for all members of result of Join operation for Pruning.
- Now, the transactions in D are scanned in order to determine L3, consisting of those candidates 3-itemsets in C3having minimum support.

Step 4: Generating 4-itemset Frequent Pattern

- The algorithm uses L3 *Join*L3to generate a candidate set of 4-itemsets, C4. Although the join results in {{I1, I2, I3, I5}}, this itemset is pruned since its subset {{I2, I3, I5}}is not frequent.
- Thus, C4= φ, and algorithm terminates, having found all of the frequent items. This completes our Apriori Algorithm.
- What's Next ?
- These frequent itemsets will be used to generate strong association rules(where strong association rules satisfy both minimum support & minimum confidence).

Step 5: Generating Association Rules from Frequent Itemsets

- Procedure:

  •For each frequent itemset ***"l"***,generate all nonempty subsets of ***l.***

  • For every nonempty subset ***s***of ***l***, output the rule **"s □ (l-s)"**if **support_count(l) / support_count(s) >= min_conf**where min_confis minimum confidence threshold.

Step 5: Generating Association Rules from Frequent Itemsets (cont.)

- Back To Example:

- We had L = {{I1}, {I2}, {I3}, {I4}, {I5}, {I1,I2}, {I1,I3}, {I1,I5}, {I2,I3}, {I2,I4}, {I2,I5}, {I1,I2,I3}, {I1,I2,I5}}.

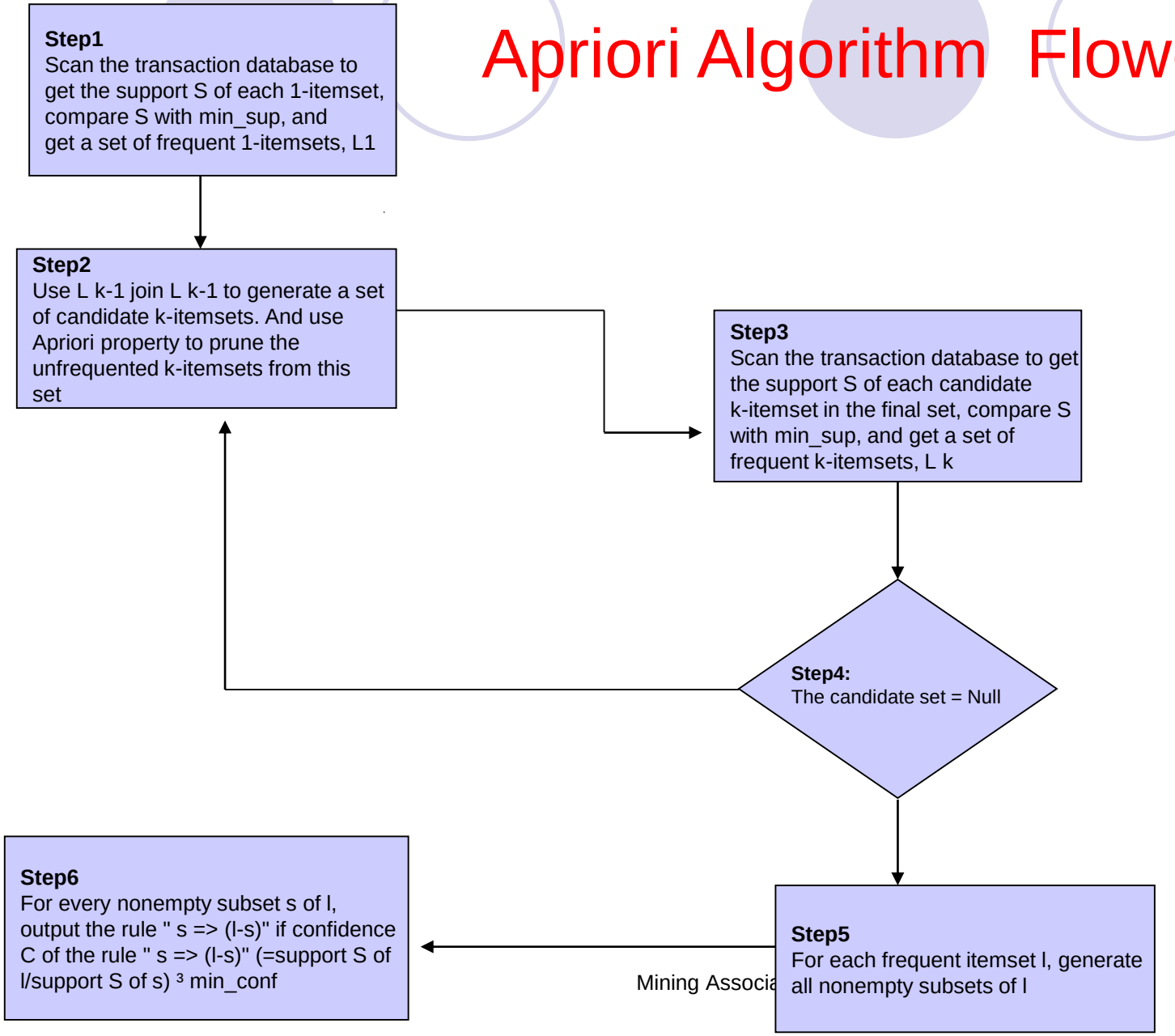
  - Lets take ***l*** = {I1,I2,I5}.

  - Its all nonempty subsets are {I1,I2}, {I1,I5}, {I2,I5}, {I1}, {I2}, {I5}.

Step 5: Generating Association Rules from Frequent Itemsets

- R4: I1 => I2 ^ I5
- Confidence = sc{I1,I2,I5}/sc{I1} = 2/6 = 33%
- R4 is Rejected.
- R5: I2 => I1 ^ I5
- Confidence = sc{I1,I2,I5}/{I2} = 2/7 = 29%
- R5 is Rejected.
- R6: I5 => I1 ^ I2
- Confidence = sc{I1,I2,I5}/ {I5} = 2/2 = 100%
- R6 is Selected.
- In this way, We have found three strong association rules.

Apriori Algorithm Flowchart



Apriori Algorithm

- Applying this algorithm on Pharmacy

- It has many interrelated transaction.

- Example:
  - My sweet baby syrup
  - Theragran
  - Rehydron – n

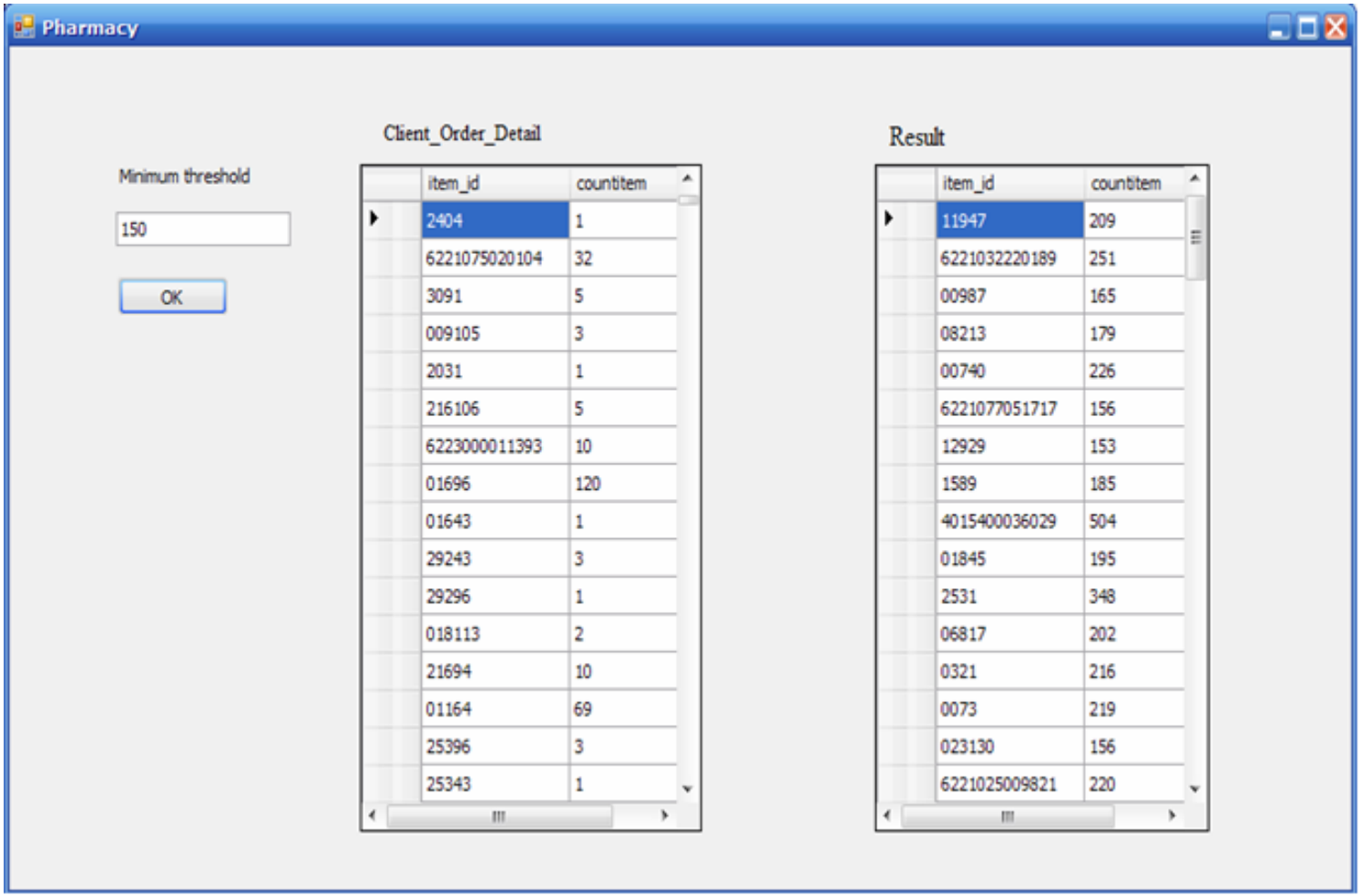
Apriori Algorithm (cont.)

- Applying this algorithm on Pharmacy

- Because it has many interrelated transaction.

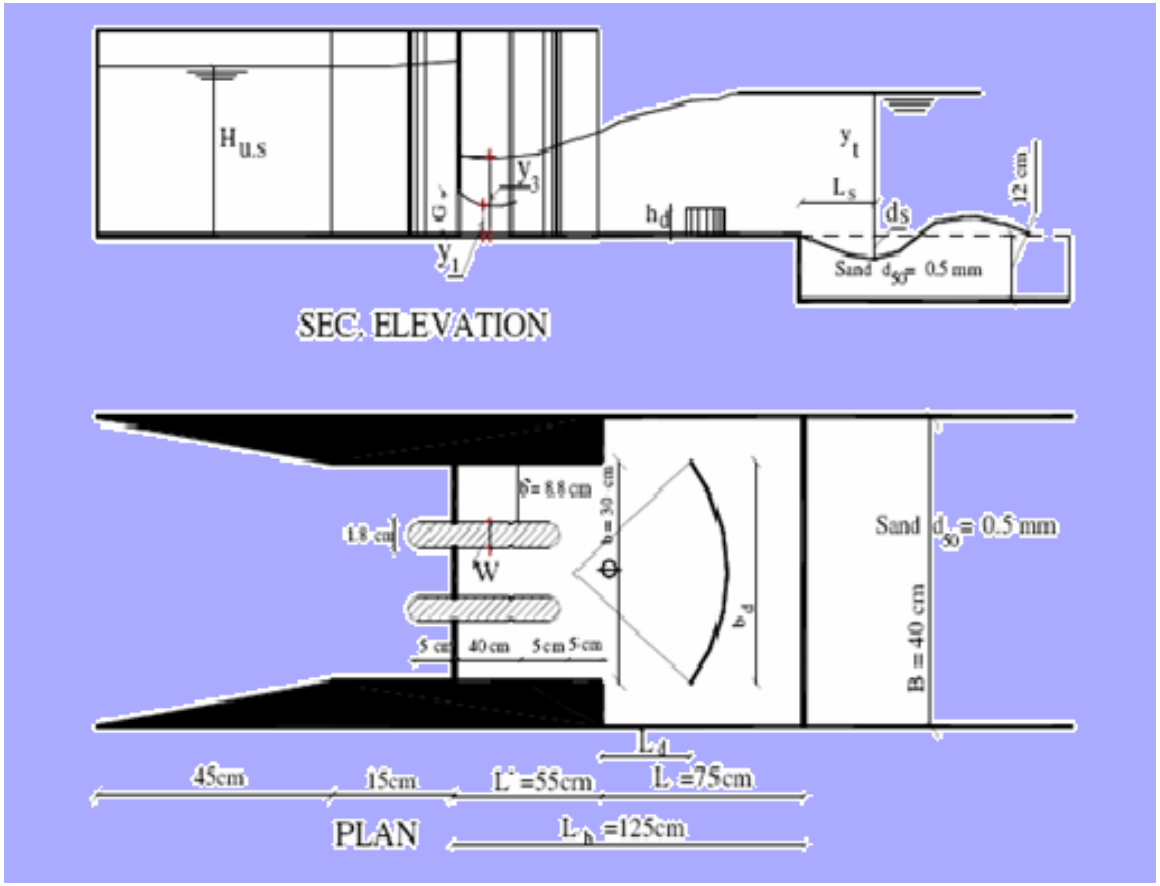
- Example:
  - Velosef500gm
  - Voltaren - 75gm
  - Brufen – syrup
  - Augmentin 25

Output Form

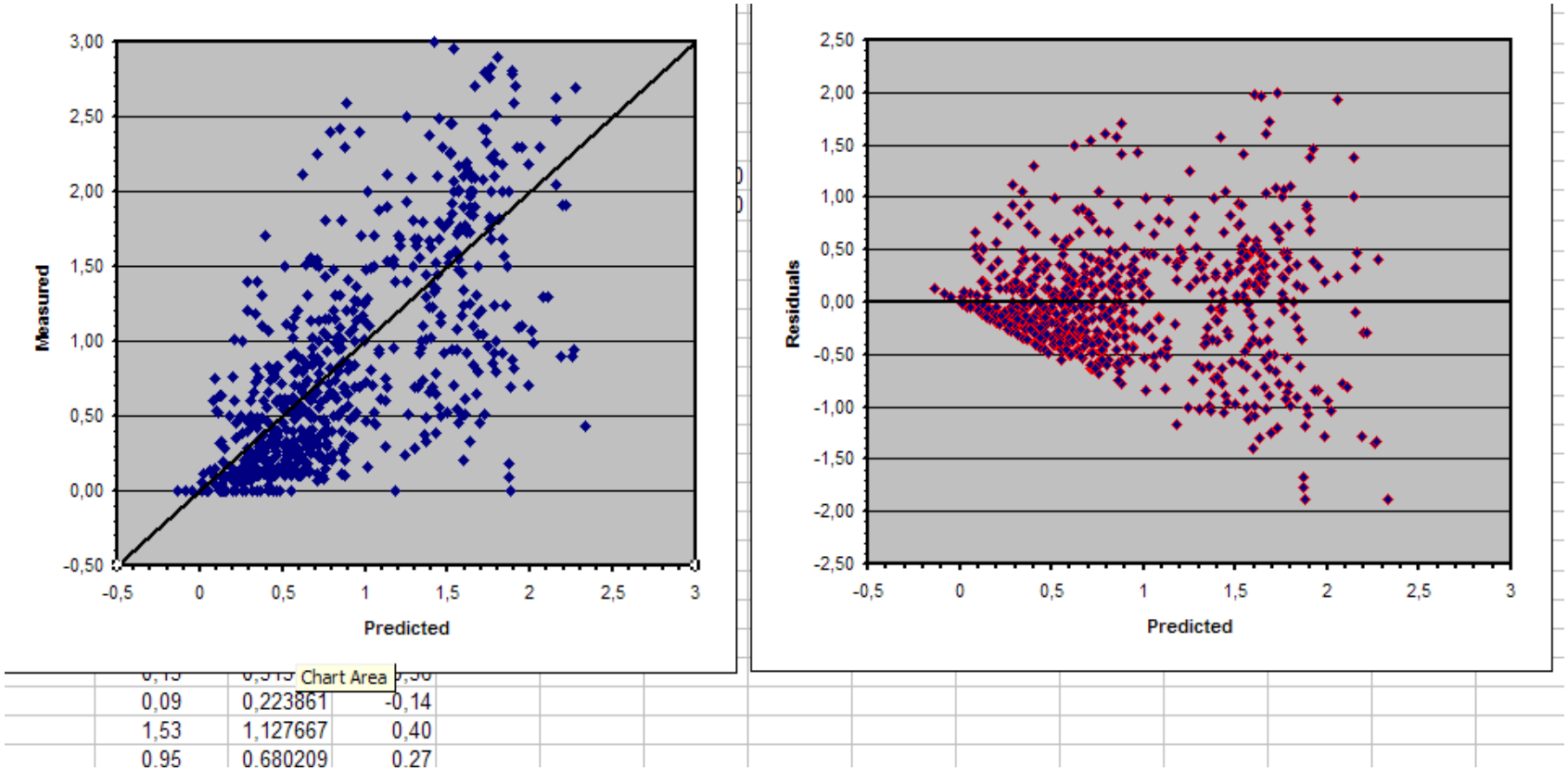

Pharmacy
Minimum threshold
150
OK
Client_Order_Detail
item_id countitem
2404 1
6221075020104 32
3091 5
009105 3
2031 1
216106 5
6223000011393 10
01696 120
01643 1
29243 3
29296 1
018113 2
21694 10
01164 69
25396 3
25343 1
Result
item_id countitem
11947 209
6221032220189 251
00987 165
08213 179
00740 226
6221077051717 156
12929 153
1589 185
4015400036029 504
01845 195
2531 348
06817 202
0321 216
0073 219
023130 156
6221025009821 220

Data Mining Using Regression Analysis

Prediction And Control Of Local Scour

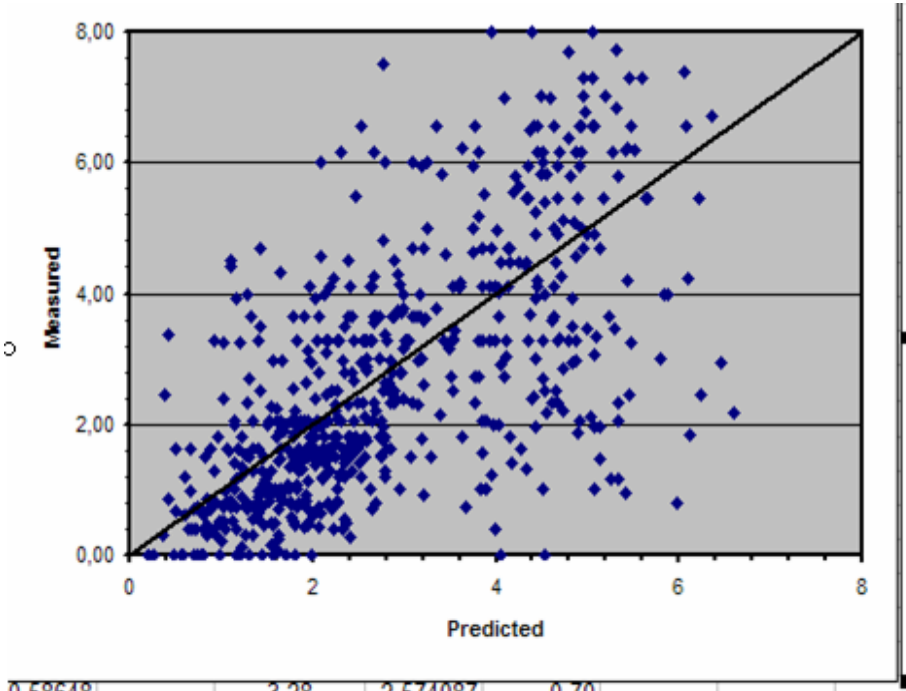


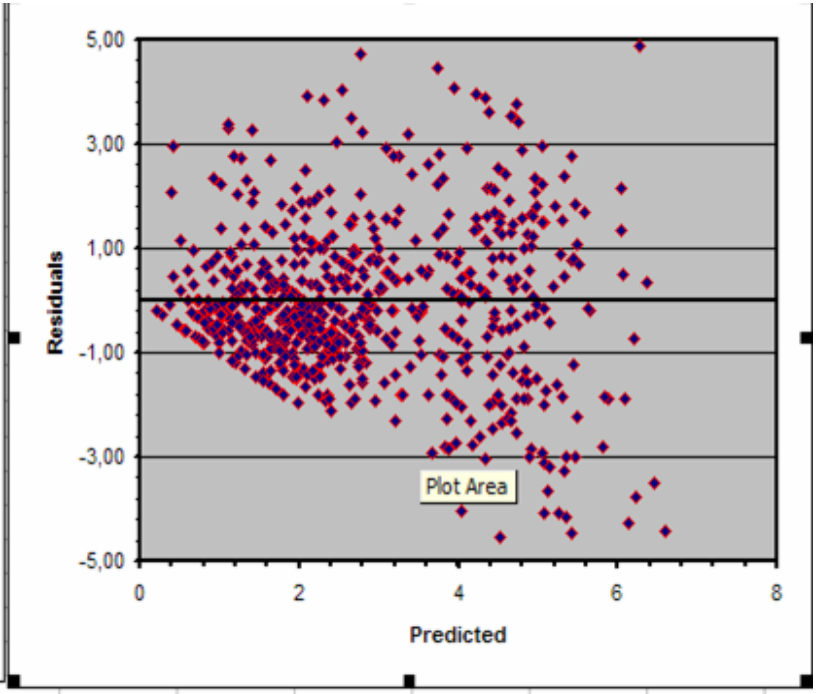
Prediction And Control Of Local Scour



This Figure : shows the Regressions Analysis that Represent Relationship between ds/y1 &Result

Prediction And Control Of Local Scour

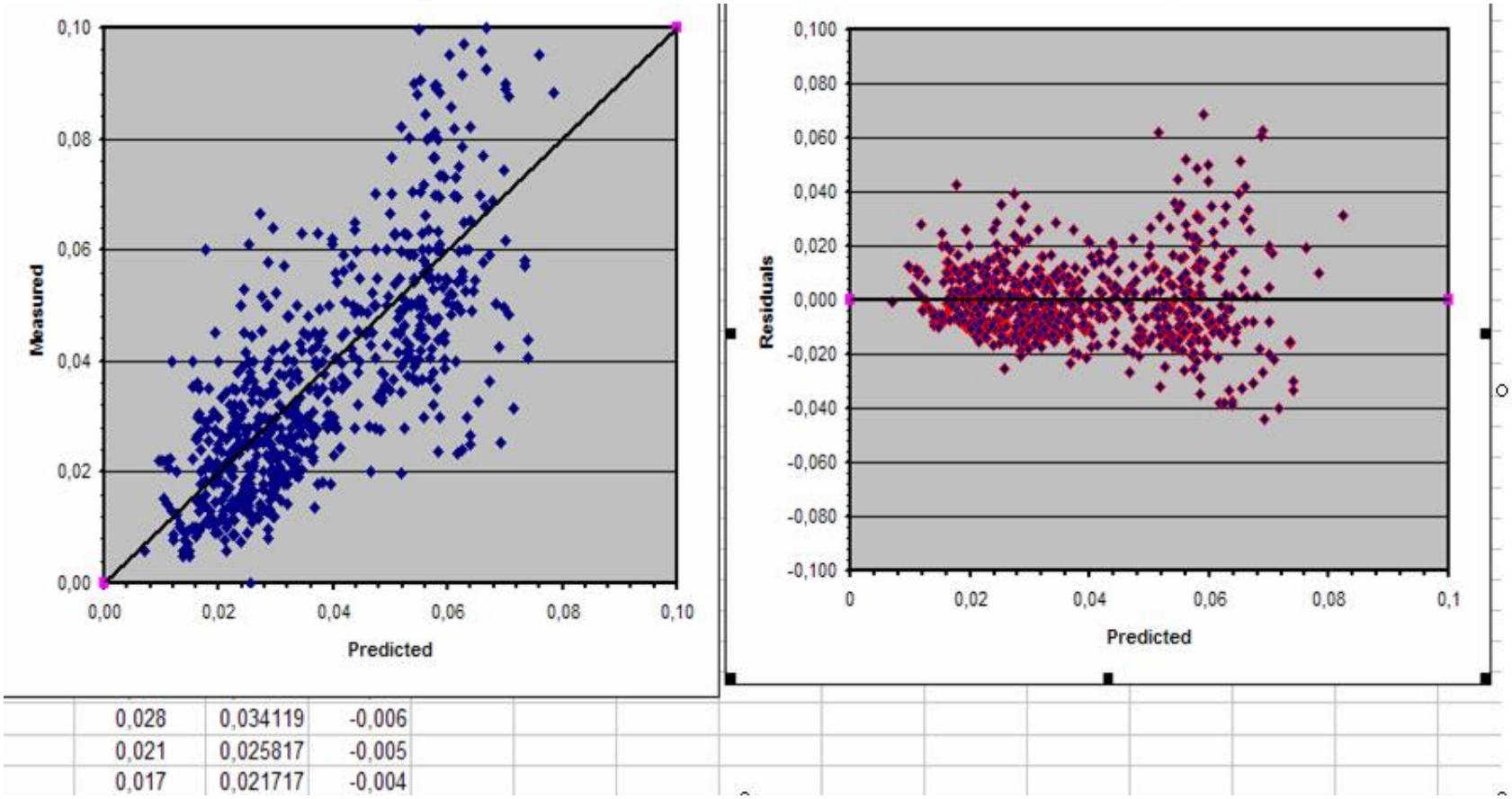




0,58648	3,28	2,574087	0,70
0,30885	3,13	1,9443	1,19
0,09686	1,60	1,075515	0,53
620381	1 75	2 762374	-1 01

This Figure :shows the Regressions Analysis that Represent Relationship between Ls/y1 & Result

Prediction And Control Of Local Scour



This Figure :shows the Regressions Analysis that Represent Relationship between Y & Result

Thank you !!!