Interesting Patterns are knowledge based [8] and are easy to understand. These are valid when tested on data with some degree of certainty and potentially useful, new or validated hunch about which the user was curious. Measures of patterns interestingness whether subjective or objective can be used to guide the patterns discovery process.

6.1 MEASURES OF PATTERNS INTERESTINGNESS

There are subjective as well as objective measures of patterns interestingness as shown in Fig. 6.1.

6.1.1 Objective Measures of Patterns Interestingness

Objective Measures of Patterns Interestingness are based on statistics. These measures specify thresholds on statistical measures of rule interestingness, such as support, confidence and lift correlations.
6.1.1.1 Support Threshold

Support represents the percentage of transactions from a database that the given rule $X \Rightarrow Y$ satisfies. This is taken to be the probability $P(X \cup Y)$, where $X \cup Y$ indicates that a transaction contains both the items $X$ and $Y$ that is the union of both $X$ and $Y$. Formally support is defined by:

$$\text{support } X \Rightarrow Y = P(X \cup Y) \quad \quad \quad (6.1)$$

6.1.1.2 Confidence Threshold

Confidence threshold assesses the degree of certainty of the discovered rule. This is taken to be the conditional probability of the rule. The probability that a transaction containing $X$ also contains $Y$. It is defined as follows:

$$\text{Confidence}(X \Rightarrow Y) = P(Y/X) \quad \quad \quad (6.2)$$

6.1.1.3 Correlation Coefficient

Coefficient of correlation is one of the most widely used statistical measures to measure the strength of relationships in two variables. Of the several mathematical methods of measuring correlation, the lift method is most widely used in practice. The coefficient of correlation is denoted by $I$ as given below:

$$I(\text{lift}) = (X \Rightarrow Y) = \frac{P(X,Y)}{P(X)P(Y)} \quad \quad \quad (6.3)$$

If $lift < 1$, then item $X$ and item $Y$ appear less frequently together in the data than expected under the assumption of conditional independence. Item $X$ and item $Y$ are said to be negatively inter dependents.

If $lift = 1$, then item $X$ and item $Y$ appear as frequently together as expected under the assumption of conditional independence.

If $lift > 1$, then item $X$ and item $Y$ appear more frequently together in the data than expected under the assumption of conditional independence. Item $X$ and item $Y$ are said to positively inter dependent.
A study was carried out in which an algorithm was developed to mine frequent patterns. Objective measures were used to discover refined patterns from large sets of frequent patterns [75].

6.1.2 Subjective Measures of Patterns Interestingness

Although objective interestingness measures facilitate identifying interesting patterns, they are ineffective unless combined with subjective measures that specify the need and interest of the user. Patterns that are expected can be interesting if they confirm a hypothesis and belief that the user wished to validate.

A study was carried out in which a two step process was used. In the first step technically interesting patterns were mined and in second step business operable domain specific filtered patterns were extracted [76].

6.2 CONSTRAINTS BASED INTELLIGENT DATA MINING MECHANISM

Constraints based Intelligent Data Mining Mechanism (IDMM) is being proposed to help the users to find relevant and valuable information. The system consists of four modules: User Input Module, Dialog Management, Inference Engine and Data Repository as shown in Fig. 6.2.

The mechanism can be used in various applications such as e-commerce, education, farming applications etc. In the present study the mechanism is assessed on the real world data set related to Socio-Economic conditions of Indian farmers. Dataset is collected through questionnaire from 350 farmers located in villages near Meerut city. The Mechanism can be used to guide various users associated with farming such as farmers, NGOs and government organization personnel working for the welfare of farmers and farming products. The aim of the proposed mechanism is to find the most relevant information for the satisfaction of the user to improve the farmers’ income and agricultural productivity.
Constraints based mining mechanism comprises following components:

![Diagram of Constraint-based Intelligent Data Mining Mechanism]

6.2.1. Dialogue Management

Dialogue management module facilitates users to select the given constraint and help the user to decide how and what type of constraints should be selected to mine the relevant information. For instance, if farmer uses the system to gain the knowledge of innovative techniques, Mechanism displays the list of constraints as shown in Fig. 6.3. Option 3 displays the dimension level constraints. System further displays the list of attributes shown in Fig. 6.4. to allow the farmer to selects the attributes. The dialogue management component is responsible for making dialogues with the users presenting the queries to the users and processing the users’ responses to propagate the further queries to the users and also input the knowledge gained from the responses to the next inference module. For instance if farmer responded association analysis, next question will be propagated to the farmer to select the IDIV/DV in which the user need the associations.
6.2.2 User Input Module

User input module provides the user friendly input/output interface and allows the user to interact with the system.

6.2.3 Inference Engine

The Inference engine is responsible for doing the mapping between dialogue management module and the IDIV/DV necessary to discover the desired patterns.

![User Interface to input constraints]

**Fig. 6.3 List of Constraints**

Inference engine accept the input from the dialogue management module which is free from all uncertainty and noise and do the mappings between user requested pattern and data sets stored in the repository.
6.2.4. Data Repository

Data repository consisting Socio-Economic factors related to Indian Farmers is used to assess the effectiveness of constraints based algorithm. Some factors are independent factors which are inter disciplined named as Inter Disciplined Independent Variables (IDIV) while dependent factors are called Dependent Variable (DV)

6.2.5 Inter Disciplined Independent Variables (IDIV)

Present study introducing the new type of variables named Inter Disciplined Independent Variables (IDIV). IDIV are those variables which affect other variables with their presence. In real life applications such as medical, agriculture, education, sale purchase and many more, certain symptoms, behavior, performance and practices depends upon various inter related factors. For instance, in medical a specific disease occurred due to various inter related factors. Different Socio-Economic conditions are responsible for the academic performance of the students. Similarly, in the field of agriculture farming practices affected from various Socio-Economic conditions of farmers. Such distinguished...
factors not only affect the dependent activity but also affect other factors and also get affected from the presence of other factors. Such types of variables are named as Inter Disciplined Independent Variables. Advantage of introducing IDIV is that these variables can be assigned threshold values which help to make algorithm more general and also help to analyze the results.

Data repository contains nine IDIV and one DV. These nine IDIV are identified as important determinants by many researchers [17, 68]. All nine IDIV are assigned four threshold values from 1 to 4. Age is an IDIV represented by IDIV\textsubscript{1}. In the survey those farmers were found below or equal to 25 years, threshold value 1 is assigned. Farmers found more than 25 years and below 50 years or equal to 50 years, threshold value 3 is assigned. Farmers found more than 50 years and below 75 years or equal to 75 years, threshold value 3 is assigned. Farmers found above 75 years or equal to 75 years, threshold value 4 is assigned.

Similarly all the Socio-Economic factors are assigned threshold values as described in section 3.4 of chapter 3.

The complete list of IDIV and threshold values is shown in Table 3.3.

DV stored in the database is income. Farmers’ income per annum in thousand is stored in repository. It is assumed that income is affected by all the IDIV resides in the repository.

6.3 **ALGORITHMS**

To assess the effectiveness of the mechanism two algorithms: algorithm-7 and algorithm-8 are being developed. Algorithm-7 is shown in the Fig. 6.5. It computes the results without using the user input constraints: measures of interestingness support and confidence. User inputs IDIV/DV are considered to find the interesting patterns. The algorithm-8 is shown in Fig. 6.6. It is developed to compute the results with user input constraints. For algorithm-8 user input two threshold values: support and confidence in addition to IDIV/DV are considered.
Algorithm-7: Without Constraint Based Data Mining
Input: IDIV₁, IDIV₂, and DV
IDIV₁ and IDIV₂: IDIV to be used to discover relationships between them
DV: Dependent Variable
Output: IDIV₁, IDIV₂, Support, Confidence, Average Income

Begin
{
while(not eof())
{
for(i=1;i<=4;i++)
{
    temp=1;
    do
    {
        if (IDIV₁ == i && IDIV₂ == temp)
        {
            c2++; // support factor
            inc=inc+income;
        }
        confidence=c2*100/c1
        averageincome=inc/c2;
        display i, temp, c2, confidence, averageincome;
        temp++;
    }
while(temp <=4);
}
}
end.

Fig. 6.5 Algorithm-7 to Mine Association Rules Without Input Constraints
6.3.2 Constraints Based Data Mining: Algorithm-8

Algorithm-8: Constraints Based Data Mining
Input: IDIV\(_1\), IDIV\(_2\) and DV
IDIV\(_1\) and IDIV\(_2\): IDIV to be used to discover relationships between them
DV: Dependent Variable
Output: IDIV\(_1\), IDIV\(_2\), Support, Confidence, Average Income

Begin
{

while(not eof())
{
for(i=1;i<=4;i++)
{
    temp=1;
    do
    {
        if (IDIV\(_1\)==i && IDIV\(_2\)==temp)
        {
            c2++; // support factor
            inc=inc+income;
        }
        confidence=c2*100/c1
        averageincome=inc/c2;
        if (c2 >= userSupport and confidence >= userConfidence)
            display i, temp, c2, confidence, averageincome;
            temp++;
    }
    while(temp <= 4);
}
}
}end.

Fig. 6.6 Algorithm-8 to Mine Association Rules With Input Constraints
Detailed Explanation of Algorithm-7 is as follows:

Database has been scanned and items were discovered that match with the user given IDIV\textsubscript{1} and IDIV\textsubscript{2}. It finds the average income of the farmers where match occurred and counts the matched association rules to compute the support threshold. Algorithm also counts the matched IDIV\textsubscript{1} and IDIV\textsubscript{2} separately to compute the confidence factor. Algorithm-7 display all the matched IDIV\textsubscript{1}, IDIV\textsubscript{2}, Average Income, and computed support and confidence thresholds without considering the user specified support and confidence thresholds. In the given example user input IDIV\textsubscript{1}: Education and IDIV\textsubscript{2}: Extension Services are considered to discover the interesting patterns. Discovered items with support and confidence thresholds are represented in the Table 6.1.

Detailed Explanation of Algorithm-8 is as follows:

Database is scanned and items are found that match with the user specified IDIV\textsubscript{1} and IDIV\textsubscript{2}. The algorithm computes the average income of the farmer where a match is found and counts the matched association rules to compute the support threshold. It counts the matched IDIV\textsubscript{1} and IDIV\textsubscript{2} separately to compute the confidence thresholds. Algorithm-8 compares computed support and confidence thresholds with the user input support and confidence thresholds and display only those items which satisfy the user specified support and confidence thresholds.

### 6.3.3 Comparison of Algorithm-7 and Algorithm-8

Algorithm-7 displays all the discovered patterns. Among these patterns some patterns are useless for decision analysis. Algorithm-8 discovered all the patterns where match found but display only those patterns where computed support and confidence thresholds satisfy the user specified support and confidence thresholds.

**Example 6.1:** User is asked to select the constraint from the displayed window as shown in Fig.6.3. If user selects the support 4% and confidence 15%, then the system display more valuable information in comparison to the condition when
user search without giving percentage of support and confidence. System represents 8 results when percentage of support and confidence is given as shown in the Table 6.2 and 14 results when this information is absent as shown in the Table 6.1.

6.4 ANALYSIS OF RESULTS

The mechanism is tested on the data set stored in farmers’ repository consisting nine IDIV and one DV: income. IDIV<sub>2</sub>: education and IDIV<sub>6</sub>: extension-services are used to assess the performance of constraints based data mining using the algorithms. Table 6.1 display 16 results. Table 6.2 output only 8 results.

Table 6.1: 16 Results without user Input Support and Confidence Thresholds

<table>
<thead>
<tr>
<th>Education</th>
<th>Extension-Services</th>
<th>Income(Th.)</th>
<th>Support (%)</th>
<th>Confidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>323</td>
<td>4</td>
<td>88</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>250</td>
<td>.006</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>281</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>256</td>
<td>8</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>239</td>
<td>9</td>
<td>41</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>300</td>
<td>.003</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>214</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>178</td>
<td>5</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>172</td>
<td>12</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>185</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>123</td>
<td>.92</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>180</td>
<td>19</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>148</td>
<td>16</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>292</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Algorithm-7 outputs interesting and uninteresting patterns both, which consume more time, more space and create confusion in decision making. Algorithm-8 displays only valuable and interesting patterns and avoids uninteresting patterns. Results conclude that constraints based mining algorithm output interesting patterns and more satisfying results. Pictorial representations of results are depicted in the Fig. 6.7.

**Table 6.2:8 Results with Support is 4% and Confidence 15%**

<table>
<thead>
<tr>
<th>Education</th>
<th>Extension-Services</th>
<th>Income(Th.)</th>
<th>Support (%)</th>
<th>Confidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>323</td>
<td>4</td>
<td>88</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>281</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>256</td>
<td>8</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>239</td>
<td>9</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>178</td>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>172</td>
<td>12</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>180</td>
<td>19</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>148</td>
<td>16</td>
<td>43</td>
</tr>
</tbody>
</table>

**Fig. 6.7 With Constraints v/s Without Constraints Data Mining Patterns**
6.5 VALIDATION OF RESULTS

Practiced algorithms of data mining in the state of the art might not suitable to find complex hidden patterns such as mutual relationships, multi dimensional relationships, multi level and cross level relationships among items from large repository. Moreover existing algorithms can be applicable on transaction database only.

A very influential association rule mining algorithm, Approri [24] has been developed for rule mining in large transaction database. Many other algorithms developed are derivatives or extensions of this algorithm. A major step forward in improving the performance of all these algorithms are the introduction of compact data structure referred to as frequent pattern tree or F.P. tree given by J. Han et al. [26]. All these algorithms allowed discovering single dimensional frequent patterns from transaction database. Most business organization store data in relational database and they need to extract complex hidden multi dimensional and interesting patterns to take valuable decisions. Moreover business decision makers require only relevant patterns that focus their needs. Existing data mining algorithms extracts both relevant and irrelevant large number of patterns. If one is interested to mine relevant patterns user assistance in the form of constraints is required. Present research study uses relational database to store Socio-Economic determinants of farmers and allows user to identify complex hidden valuable patterns under constraints based framework.

Literature shows that there is a big gap between the existing data mining applications and the interest of industrialists, scientists and experts. Traditional studies uses statistical measures to find the relationship among various items stored in the repository. These results are not as explanatory and interpretive as the results obtained by association rule mining algorithms developed in the present work. Malathesh et al. [21] find the contribution of various Socio-Economic Factors on extension use efficiency under crop farming system using statistical
techniques. The same statistics is applied on the sample data set shown in Table 6.3 extracted from the repository developed in the present work and obtained results are compared with the results obtained from association rule mining algorithm developed in the present work. Table 6.4 indicates the results obtained from statistical techniques and Table 6.5 and Table 6.6 indicate the results obtained from association rule mining algorithm developed in the proposed work.

Table 6.3: Sample Data Extracted From Database Containing Threshold Values of Socio-Economic Factors of Farmers

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>IDIV₂ (Ed)</th>
<th>IDIV₉ (Ex)</th>
<th>DV Ic(Th.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>400</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>500</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>499</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>600</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>560</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>500</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>550</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td>590</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
<td>650</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>3</td>
<td>600</td>
</tr>
</tbody>
</table>

Table 6.4: Results obtained using Statistical Technique

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variables</th>
<th>Regression Coefficient</th>
<th>SE of Regression Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>Extension use efficiency</td>
<td>.812</td>
<td>.509</td>
<td>.3434</td>
</tr>
</tbody>
</table>

Results shown in the Table 6.4 indicate that extension use efficiency is affected by the educational background of the farmers. Same sample of ten transactions is used by the association rule mining algorithm-3 developed in the present work and
results are indicated in the Table 6.5 and Table 6.6. User input support threshold 20% and confidence threshold 50% are considered.

**Table 6.5: Results obtained from Association Rule Mining Algorithm-3, s=20% & c=50%**

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>IDIV&lt;sub&gt;2&lt;/sub&gt;:Ed</th>
<th>IDIV&lt;sub&gt;7&lt;/sub&gt;:Ex</th>
<th>DV: Income (Th.)</th>
<th>Support %</th>
<th>Confidence %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>400</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>500</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>549</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>600</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>600</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>525</td>
<td>20</td>
<td>66</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>650</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>3</td>
<td>600</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

**Table 6.6: Strong Rules obtained from Association Rule Mining Algorithm**

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>IDIV&lt;sub&gt;2&lt;/sub&gt;:Ed</th>
<th>IDIV&lt;sub&gt;7&lt;/sub&gt;:Es</th>
<th>DV: Income</th>
<th>Support %</th>
<th>Confidence %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>549</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>525</td>
<td>20</td>
<td>66</td>
</tr>
</tbody>
</table>

Results obtained by Association Rule Mining algorithm are more explanatory. Results provide the threshold value of support and confidence for each transaction. Support indicates the frequency of the pattern in database and confidence indicates the probability of occurring the item 2 when item 1 is present. Confidence indicates the strength of the rule. Strong rule identifies those rules which satisfy user input support and confidence thresholds. It concludes that the results obtained by proposed association rule mining algorithms are far better than the results obtained by statistical measures.
6.6 SUMMARY

The constraints based data mining technology has generated new dimensions and opportunities of customization. Constraints limit the Search space and help computational statements to specify where to begin. How to calculate the path to descent and when to terminate the search. In this chapter, issues related to customization are discussed in the light of constraints based mining of data from large data set. New types of variables named inter disciplined independent variables are discussed and threshold values are assigned to these variables to make the results more valuable. The chapter discusses user specified interactive explorative constraints based data mining that is assessed on real world data set related to Socio-Economic conditions of farmers and resulted that constraints based data mining produce more valuable, concise and concrete results as compare to non constraints based data mining.