CHAPTER- 3

METHODOLOGY OF THE SYSTEM DESIGN

The need for the modern corporate performance management unifies the so-called “Big Three” processes of planning, budgeting, and forecasting in sustaining accuracy, consistency, and transparency. The ‘efficient and effective’ integration of these processes involves exploring detailed information on predictive as well as on real time cases. Data mining, popularly known as ‘Knowledge Discovery in Databases’ [39], is the right process of discovering inherent valuable patterns of data in databases to aid in meaningful decision-making [29]. Data mining can provide the potential ‘financial’ and ‘competitive’ advantage to an organization by exploring the large data warehouses.

This target-oriented approach warrants databases of either operational or informational data warehouse that make the corporate data readily accessible and usable for managerial decision-making. From such data warehouses the complementary technique of data mining is made with its extensive applications in almost all fields of business problems. Further, data mining has relevant capability to build predictive rather than retrospective models [30].
3.1. DATA MINING TAXONAMY

The various phases of data mining [31] involve the selection of data for analysis from operational databases. The next phase is cleaning the data which is referred to as preprocessing of data. This cleaning operation removes the inconsistencies and discrepancies among the data extracted from the data warehouse/databases [32]. The dataset is now analyzed to identify patterns. The model is validated applying the experimental data to discover knowledge. If the model works well it should be possible to translate the concerned model into an action oriented process to achieve the targeted reach.

Such a model to discover business processes satisfying all the above mentioned conditions becomes a model of Business Knowledge. It is noted that the above steps are iterative until the best quality business knowledge [33] is obtained.

From the above discussion, it is understood that data mining improves the ability of identifying the hidden patterns in data. Historic data either from a single or from multi source operational databases might be heterogeneous in nature. As the predictive accuracy of data mining strongly influences business solutions, the quality of data is to be improved through conscious elimination process of noisy data and also by handling the missing data [34].
Thus the predictive accuracy of data mining technique strongly influences business solutions. To coordinate with enterprise type of business applications, the data mining should integrate with other decision support systems (DSS) and database management systems (DBMS) in the organization. This desirable technique of data mining helps in binding the remote databases which upstream the idea of e-business among the enterprises world wide.

3.2. BUSINESS PROCESS MINING

Innovative e-business design creates prompt, desired business values satisfying the customer priorities of today and tomorrow. Thus targeting the wish of the satisfied customer, the design of the e-business starting from the raw material supplier to the retail marketer should engineer all the business traits along with their workflows to make every action a business process function [35]. At this stage it is necessary to speak about the legacy systems which are maintained for a number of years. A conventional legacy system, i.e., the absence of any information system domain, may not provide an efficient business solution either to the entire enterprise level or to any of its functional levels. For instance, the inefficient legacy system starts with extensible time of operations which is the base for the uncontrolled inventory. Hence, it is essential to have an ‘Intelligent’ system, i.e., IS
oriented system, to fetch the valuable inherent data based on queries of the domain user.

The practical adoption of intelligent business systems focuses reduction of time consumption for process mining and functions, resulting in improved turnovers and avoids bottleneck moments in the industry [36]. In simple words, IS oriented system is expected to be a ‘decision enabler’ rather than a simple ‘data store manager’ [81]. Pursuing this line, the proposed hybrid system is designed to work as an ‘Intelligent Information Processor’ even without any support of ‘Enterprise Packages’.

Emerging technologies easily manage the growing challenges [37]. For a company to be successful, it should no longer add values but should invent values. Exploring business values to any particular function is oriented in the relevance of the data leading to that function. This is the right stage where the techniques of data mining are used to accomplish the action oriented digging. The supply of such polished patterns of data acts as inputs for the function of the packages like supply-chain management, customer-relationship management, enterprise-resource planning, and selling-chain management.

Such information processes on one to one merger with each other improve the functions of the enterprise in toto and when
effectively routed on the functional channel of business, unmark actions like 'commit', 'schedule', 'make', and 'commitment time', referenced to the independent dimensional domain – 'the time'.

3.3. OVERVIEW OF THE HYBRID DESIGN

Previous discussions explicitly mention that a better technology performance results in a lesser time for process execution with an IS
oriented system compared with the conventional legacy system. It is for this reason the methodology attempts re-engineering an IS oriented system by applying the present day trend of computational fusion technique to positively shift the technology so that the quality of the enterprise takes the stand of competitive advantage.

On making the decision of incorporating IS over an enterprise, the next stage of the move is discovering knowledge from the historic databases in the applicatory fields of business [38]. Historic databases have inherent values and now begin our process of polishing the data. Polishing of data is ventured using rule induction technique and analysis of data is made using the proposed data mining algorithm to satisfy the interaction of the strategic development of the application. Validating the model is done employing the proposed rule based algorithm on a multi layer feed forward neural network. This whole some technique makes the information system a fusion or hybrid model of RI, data mining, and ANN.

Data mining techniques should be always directed for making confirmatory analysis. Fully directed techniques require a-priori specifications like inputs, outputs, and models. Small models while easy to comprehend may simply be inadequate in making depictions of the active relationships whereas large or complex or sometimes hybrid
models even though they suffer from difficulties in understanding provide improved predictive capabilities. Similar statements can be also provided in case of undirected inputs. The model predicts only the end points of a continuum of semi-directed approaches and hence the results might be unbelievable.

Thus keeping the objective of information systems perspective on business traits the most meaningful and pattern detective system design namely Rule Induction set in with Fuzzy database is taken for knowledge discovery applying the core concept of Apriori algorithm worked on a spanning tree. The results of the data mining module are then trained and tested using a multi-layer Artificial Neural Network applying the rule based algorithm for business process mining.

On integrating all the above said perspectives the mining automation underlying the processes lifts up the outcome of the performed business cases.

This perspective on automation stand incorporates rule induction technique on databases to categorize them homogeneously. These categorized databases are converted into fuzzy databases using fuzzy membership functions. Ranges of fuzzy transforms are taken because ‘finer’ predictions will be possible on a wider range when they are worked on data sets. At this stage the design of the system incorporate
Apriori algorithm on the fuzzy transformed tree structured data sets. This procedure is hereafter termed as Adaptive Fuzzy Apriori-Tree search algorithm. This data mining algorithm together with the ANN trained rule generator mine knowledge and make predictions.

Artificial Neural Networks (ANN) belongs to the family of learning-by-example paradigm in which problem solving knowledge is automatically generated according to actual examples. ANN technology offers extensive support in summarizing, organizing and classifying data. It acquires a higher degree of predictive accuracy on operational data sets. The learning and the recall procedures of ANN replicates the learning and the recalling processes of the human brain.

The essential practical features of ANN are parallel processing, error tolerance, recall memory, and optimization of solutions. Integrated Fuzzy - ANN model has been used now-a-days in prominent areas [40] of application. In the wide range of business applications the system suggests the hybrid approach [81] of Rule Induction with the integrated Fuzzy-cum-ANN knowledge discovery and decision making to produce optimized parameters in business process mining. This integrated Fuzzy-cum-ANN system employs the fuzzy transformed analysis approach of transactional databases using the proposed algorithms.
Figure – 3 Phases involved in making predictions using the hybrid model
Figure-3 shows the different phases of knowledge discovery and decision-making compartments of the suggested model. The various stages of the suggested hybrid model can be explained as follows.

(i) The input to the model is a raw transactional database which may be even operational.

(ii) Based on this master database, a cluster of sub-databases are to be extracted by applying the RI technique, i.e., a technique focused on a specific (some) condition(s). The extracted cluster of sub-databases may resemble homogeneous set due to the rule condition specified by RI and hence, improves efficiency in providing valuable information. These sub-databases contain the crisp form of classified data.

(iii) This step involves the fuzzy transformation of data from the crisp sub-databases by using fuzzy transform functions. In other words, on integrating the fuzzy engine to the crisp data, it is possible to transform the crisp data into the fuzzy data illustrating quantitative business
processes such as finance, customer queries and transactions, marketing transactions, etc.

(iv) This is the third stage in which Adaptive Fuzzy Apriori-Tree search algorithm is applied to distinguish potential itemsets of transactions on the fuzzy database. The core phase of the proposed algorithm implemented on the database generates the all samples itemsets transactions in the database. Now the scalar cardinality for the nonzero itemsets is calculated to segregate the mined data. To make the rules more interesting and precise the appropriate threshold value is set for support.

(V) The present stage applies the ANN rule generator on the mined itemsets, to provide all possible predictions.

The hyper interacted predictions lead to an efficient solution suggested by expert professionals. Thus this hybrid model works for knowledge discovery and hence supports decision-making. The application of trained interrogations through this model results in better expertise, resulting in a decision support system.
3.4. DYNAMICS OF RULE INDUCTION

The increasing use of a very large number of databases and data warehouses that mine useful information and knowledge which lead to effective decision-making gains more interest in almost all real time and operational transactions. Apart from the conventional type data mining using binary values, the real world quantitative values are supplied on which the system performs the proposed design.

The analogy of Rule Induction (RI) with the most powerful ‘if–then–else’ reasoning which reduces the complexity of data analysis by the monotone way (many items satisfying the same rule) of classifying the databases which is popularly called ‘hashing’ using conditions [41] is employed in this fusion model as the prime stage. The methodology is worked on the master database to hash it, to prune it, and to clean it. At this stage preprocessing of data is made so that the missing, least preferred, and the dominant data sets are identified and segregated. As a result the homogeneous data sets are generated which are used for the system design.

RI is one such intelligent business process of analyzing the practical applications like customer profiling, product analysis, company analysis and so on. In the proposed system design, the data set transactions are deep sorted by applying rule based conditions. The
approach helps to comprehend and validate the framed design. The tactics of deep sorting the databases generate interesting and crisp appearances on to the specific conditions.

The dynamics of inducting rules in the system design thus helps in segregating databases homogeneously from which the analysis of knowledge digging is made. Suggestions pruned and based on such illustration develop needful decisions [42, 43].

3.5. R-I ON THE DESIGN ARCHITECTURE

The key question of applying RI is, to identify which of the attributes is the most useful to determine the conclusion of the rules. Whichever it is, that attribute ought to be the first one in decision making. Information theory provides the simplest answer that is ‘RI is providing a meaning for the notion fixed making the condition a most useful determiner’. In the proposed work the heterogeneous master database is a collection of 189 data sets.

These datasets were identified as data regarding the customers who invest in the stock market intentionally either for the sake of Investment or for the sake of Business. At this juncture it is so decided to analyze the risk factors involved in the action of the customers
separately. To achieve this, the model is designed to segregate the customer behavior on the type of investment.

It is very appropriate that rule induction technique is applied on the heterogeneous database to hash it into collection of homogeneous databases satisfying the applied condition as the rule. Thus the attribute chosen is the ‘type of investment’ worked on the condition whether type of investment is for business sake or type of investment is for investment sake and the master database has been hashed (sub-divided) into two databases one called the ‘BUSINESS’ database and the other known as the ‘INVESTMENT’ database. The rule inducted pruned database now consists of

Number of records in the Business database \(-\) 71

and Number of records in the Investment database \(-\) 118.

This RI over the master data sets helps in dealing with the most diversified approach which will lead to wider and needy decisions making the system an able management information system. These homogeneous databases are now handed over to the next stage of the model for knowledge discovery.
3.6. KNOWLEDGE DISCOVERY MODULE

3.6.1. FUZZY TRANSFORMS

The knowledge discovery model has major components which include a knowledge discovery interface, a fuzzy engine to generate a fuzzy database and the modified Apriori algorithm that has been derived from the all samples cluster. The knowledge discovery interface is an interface for inputting the raw data into the model. This data is called the crisp data. The main objective of the fuzzy engine is to transform the crisp data into fuzzy data using the fuzzy membership functions [44].

The major tasks of the fuzzy engine include defining the terms of fields, arranging the membership functions, and transforming the crisp data into fuzzy data. Under most conditions, there are three to seven functions used to present the strength of the fuzzy data. Out of these the four types of standard membership functions widely employed are Z-type, A-type, Π-type, and S-type in which the A-type (triangle type) is the most versatile type for applications.

Fuzzy values have been extended to handle the concept of partial truth and truth values between "completely true" and "completely false". As its name suggests, it is the logic underlying modes of
reasoning which are approximate rather than exact. The importance of fuzzy derives from the fact that most modes of quantitative reasoning are approximate in nature [45, 46].

3.6.2. GENERATING FUZZY DATA BASE IN THE SYSTEM DESIGN

This module involves the fuzzy transformation of the crisp sub-databases by using fuzzy membership functions. Why do fuzzy transforms are employed at this stage is a valued question. The answer to this question is theories about fuzzy transforms are highly valuable in predicting knowledge when worked on real world linguistic or highly subjective or humanistic datasets. The proposed methodology handles quantitative values in a fuzzy database. Most of the data mining works handle values in binary converted format. This module deals with the ideology of training the real time quantitative values and for such values the fuzzy approach could be suggested, which can trail a smooth transition of continuum of values with appreciable boundaries. The master database considered as a real time application to work on the proposed hybrid model deals with financial implications like the amount of investment in the stock market, recurring income, and property asset and also the age of the customer making investment. Considering such variables like money transactions and age of the
investor it is obvious that such data sets are highly subjective
depending on the social and environmental circumstances. Hence it is
decided to apply fuzzy transforms to the crisp data fetched from the
interface and that has been made by the fuzzy engine in the design
architecture.

The next valued question is that for transforming the crisp data
into fuzzy data, what kind of transformation has to be adopted. That is,
which type of transform membership function has to be applied? At
this stage the system takes into the selection of ‘Triangular’
membership functions. This membership function has been selected for
this work because it is a proven fact that triangular membership
functions are highly versatile on real world applications. Thus for the
supplied quantitative training data the logic of fuzzy transforms
generates the membership values which are collected as fuzzy values
and stored in a database very popularly called as the fuzzy database.

At this stage the system has to be designed to align the triangular
membership functions for use of the model. For this to be done, it is to
be considered how many membership functions are to be included in
the fuzzy engine and what the limits of continuum to build the
functions are. It is analyzed to have three membership functions named
‘Low’, ‘Medium’, and ‘High’ assigned to the decision variables age,
amount of investment, term of investment, recurring income, and property asset value. The optimum limits for the variables, to convert them into fuzzy is fixed scanning the master crisp data base and is scaled as shown below.

** Amount of investment, Recurring income, and Property asset value are financial deals expressed in terms of the Indian currency ‘Rupee’, Age is in ‘Years’, and Term of investment is in terms of ‘Months’.

<table>
<thead>
<tr>
<th>AGE (in years)</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18</td>
<td>30</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AMOUNT OF INVESTMENT (in terms of Rupee)</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1000</td>
<td>20000</td>
<td>80000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TERM OF INVESTMENT (in months)</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

- 43 -
RECURRING INCOME (in terms of Rupee)

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>900</td>
<td>4000</td>
<td>9000</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>10000</td>
<td>100000</td>
</tr>
</tbody>
</table>

PROPERTY ASSET VALUE (in terms of Rupee)

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25000</td>
<td>90000</td>
<td>900000</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>1000000</td>
<td>10000000</td>
</tr>
</tbody>
</table>

This optimized scaling fits into all conditions of designing fuzzy membership functions and this design is achieved applying the below mentioned formula.

If ‘A’ is a fuzzy number and only if there exists a closed interval

\[[a,b] \neq \Phi\]

such that

\[
A(x) = \begin{cases} 
1 & \text{for } x \in [a,b] \\
1(x) & \text{for } x \in (-\infty, a) \\
r(x) & \text{for } x \in (b, \infty), 
\end{cases}
\]
where

\[ l \text{ is a function from } (-\infty, a) \text{ to } [0,1] \text{ which is monotonically increasing, and such that } l(x) = 0 \text{ for } x \in (-\infty, \omega_1); \]

\[ r \text{ is a function from } (b, \infty) \text{ to } [0,1] \text{ which is monotonically decreasing, and such that } r(x) = 0 \text{ for } x \in (\omega_2, \infty). \]

Applying the prescribed analogy the designed fuzzy engine transforms the crisp data into fuzzy data and the resulting database is called the fuzzy database. As in the previous stage of the model, RI sub-divides the master database into two sub-databases the fuzzy engine now yields two fuzzy sub-databases, hereafter termed as business fuzzy database and investment fuzzy database.

3.6.3. ADAPTIVE FUZZY APRIORI (AFA) – TREE SEARCH ALGORITHM

In the proposed model the fuzzy transformations are assigned to quantitative values to extract all possible versus impossible
classifications and combinations of all degrees between the assigned intervals say \((i_1, i_2)\). Here the maximum yield point is fixed up say \(i_{\text{max}}\) and naturally by fuzzy logic \(i_{\text{max}}\) intends to the value 1. Thus all values of the fuzzy set are to be organized in between the points of minimum and maximum and this happens by assigning the membership functions. The type of function depends on the nature of the problem treatment but with regard to all and any type of problem assignment, the triangular membership function suits describing the set values very well. Hence the system utilizes the triangular membership functions for training the quantitative data. Thus the quantitative data is being converted into fuzzy data and stored in the fuzzy database. Now the application of the data mining tool is proposed on the fuzzy database.

The proposed data mining algorithm classifies and mines all possible dominant itemsets, which enables knowledge discovery. Itemsets are collection of data sets satisfying some conditions. On grounds of quantitative approach those datasets are called itemsets. The classification taxonomy is performed adapting the proposed algorithm on the fuzzy data.

The proposed Adaptive Fuzzy Apriori (AFA) - Tree Search Algorithm is illustrated as below.
The first step of the proposed algorithm essentially clusters all needed decision variable samples, hereafter termed as ‘all samples itemsets’. This approach of clustering all needed decision variables as samples and considering only the all samples itemsets for application of Apriori algorithm follows the ‘top-down’ spanning tree data structure approach in mining knowledge from the database intended to reduce time complexity.

Now the algorithm checks for nonzero values in the fuzzy database for the clustered all samples itemsets and for all nonzero combinations it fetches the minimum fuzzy values which are stored in a temporary set $C_r$. Now the scalar cardinality for the minimum nonzero values of the all samples itemsets is calculated and if this count satisfies the minimum support then the attributes of such linguistic terms are moved to $L_r$, which is the set of potential itemsets.

Starting with $r = 1$, the first iteration generates the first set of all samples itemsets (join step) and if the availability of all items in the samples itemsets is nonzero then for the minimum value the count is performed (prune step) and verified for support threshold. At this stage the algorithm indulges the condition of supply of a ‘Support threshold’ value by the expert professional. The value can be low or high, accordingly the cutting edge of knowledge discovery lies.
When \( r = 2 \), the successive all samples itemsets are generated to the cycle for which the counts are calculated independently for availability of items and verified for support.

The above step is repeated for any number of cycles of iteration until no all samples itemsets is left in making cluster or in reaching the support threshold.

Thus the algorithm is made to run on the all samples itemsets based on the theory of satisfying minimum support.

The procedure of AFA-Tree search algorithm on all samples itemsets includes the following steps.

**Step: 1**

Let \( A \) be the crisp dataset.

**Step: 2**

The crisp data \( A \) is transformed into fuzzy data using the triangular membership functions of ranges ‘low’, ‘medium’, and ‘high’ with appropriate mapping values.
Step: 3

If the membership function value of the input crisp data $A$ at the $i_{th}$ record, $j_{th}$ field and $k_{th}$ term is $\mu^{(i)}_{jk}$ then the $j_{th}$ field membership function of the input data $A$ of record ‘$i$’ is given as

$$A = \sum (\mu^{(i)}_{jk} / X_{jk})$$

where $X_{jk}$ is the $k_{th}$ fuzzy region of $A_j$.

Step: 4

Let $r = 1$.

**Join step of all samples**

Cluster all possible combinations of all needed samples of decision variables which form the all samples itemsets for knowledge discovery. The search stores the items in a temporary set $C_r$.

**Prune step of all samples**

The prune step scans for all nonzero valued itemsets and fetches the minimum value ($\mu$) of the variable of the combination of all samples itemsets and finally sums up for scalar cardinality.

Thus for a database of $n$ records the scalar cardinality called as $\text{Count}$ is
\[ \text{Count} = \sum_{i=1}^{n} \mu_i \]

**Step: 5**

If \( \text{Count} > \ell \) the minimum support threshold stores the results in \( L_r \).

**Step: 6**

Repeat the above steps for all possible combinations.

This algorithm is patent to be the Adaptive Fuzzy Algorithm (AFA)-Tree Search algorithm worked on all samples of itemsets.

The fuzzy database constructed is modeled to form a fuzzy transform matrix on which the algorithm is applied, generating the potential all samples itemsets satisfying the minimum support threshold. A 5 by 3 matrix is framed with the elements as shown below and data search over the matrix span is made from the first to the last row scanning each and every item of each column of the matrix. This is termed to be the top-down or decisive data structure approach of a spanning tree in which the data cluster starts with the root spanning to the other leaf nodes mining all probable combinations.
keeping the root nodes fixed. Pertaining to the problem which is applied in the hybrid model the decision variables say age, amount of investment (amt), term of investment (term), income, and asset form the decisive span structure are as shown below.

1. Age-low \rightarrow Age-med \rightarrow Age-high
   \rightarrow Amt-low \rightarrow Amt-med \rightarrow Amt-high
   \rightarrow Term-low \rightarrow Term-med \rightarrow Term-high
   \rightarrow Income-low \rightarrow Income-med \rightarrow Income-high
   \rightarrow Asset-low \rightarrow Asset-med \rightarrow Asset-high

2. Age-low \rightarrow Age-med \rightarrow Age-high
   \rightarrow Amt-low \rightarrow Amt-med \rightarrow Amt-high
   \rightarrow Term-low \rightarrow Term-med \rightarrow Term-high
   \rightarrow Income-low \rightarrow Income-med \rightarrow Income-high
   \rightarrow Asset-low \rightarrow Asset-med \rightarrow Asset-high
It is found that each variable has three membership functions low, medium, and high with which the mathematical model of the decision tree set up is illustrated. On such a tree, scan for combinations applying AFA-tree search algorithm yields about 27 clusters each having 9 combinations about which a wide discussion is made in chapter 5. The appearance of such combination for example resembles (agelow, amountlow, termlow, incomelow, assetlow). An important point to be noted in this analogy is redundant itemset variables like (agelow, agelow, termlow, incomelow, assetlow) or (agelow, amountlow, incomelow, incomelow, assetlow) are eliminated. If so totally about 243 combinations are framed and the algorithm scans for nonzero values for each variable and fetches the minimum for that
sample combination to make the count. If the count satisfies the minimum support the combination is termed to be ‘potential’.

3.7. DECISION MAKING MODULE

The previous module discusses about the knowledge discovery of potential investors in stock market. This module explains the decision making stage of the hybrid model which explores all possible predictions and hence suggestions for least risk investment on trade. It is achieved using a multi layer feed forward perception technique popularly known as the Artificial Neural Network (ANN) [47]. The ‘Rule-Based’ algorithm is proposed and incorporated to train the network and test the data. The rule based algorithm is suggested because most of the real world quantitative values are scalars and the proposed hybrid system is designed to handle such values. Such scalar values are found efficiently manipulated by the proposed algorithm and hence it has been utilized for run of this supervised kind ANN.

3.7.1. OVERVIEW OF ANN

An Artificial Neural Network (ANN) or commonly just Neural Network (NN) is an interconnected group of artificial neurons that uses a mathematical model or computational model for
information processing based on a connectionist approach to
computation. In most cases an ANN is an adaptive system that changes
its structure based on external or internal information that flows
through the network.

In more practical terms, neural networks are non linear statistical
data modeling tools. They can be used to model complex relationships
between inputs and outputs or to find patterns in data [48]. An
Artificial Neural Network (ANN) is an information processing
paradigm which is inspired by the biological nervous systems. The key
element of this paradigm is its novel structure such as the brain
processing information. ANN like people learns by example [49]. This
taxonomy of ANN working with the rule based algorithm makes the
proposed module a rule based valuator [50] in making suggestions.

3.7.2. IMPLICATION OF ANN IN THE HYBRID MODEL

Neural Networks, with their remarkable ability to derive
meaning from complicated or imprecise data, can be used to extract
patterns and detect trends that are too complex to be noticed by either
humans or other computer techniques. A trained Neural Network can
be thought of as an "expert" in the category of information it has been
given to analyze. This expert can then be used to provide projections
given new situations of interest and answer "if-then-else" questions. Other advantages include:

(i) Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.

(ii) Self-Organization: An ANN can create its own organization or representation of the information it receives during learning time.

(iii) Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.

(iv) Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to a corresponding degradation of performance.

Neural Networks take a different approach to problem solving from that of conventional computers. Neural Networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem [51]. The examples must be selected carefully otherwise useful time is
wasted or even worse the network might be functioning incorrectly and its operation can be unpredictable.

On the other hand, conventional computers use a cognitive approach and if anything goes wrong it is due to a software or hardware fault. Neural Networks and conventional algorithmic computers are not in competition but complement each other. There are tasks that are more suited to an algorithmic approach like arithmetic operations. But tasks like real words cases are more suited to Neural Networks. Even more, these real world cases require systems that use a combination of the two approaches (normally a conventional computer is used to supervise the Neural Network) in order to perform at maximum efficiency [52].

![Simple model of a multi layer ANN](image)

**Figure - 4 Simple model of a multi layer ANN**
The simplest model ANN used in the hybrid model is as shown in figure - 4. This model supply input variables to input neurons, hidden neurons to perform computation based on rules, and output neurons display the results. Data modeling and application of rules are made in the hidden neurons. The supervised kind of learning algorithm is employed in the system. Supervised learning incorporates an external teacher, so that each output unit is told what its desired response to input signals ought to be.

Paradigms of supervised learning include rule based error-correction learning, reinforcement learning and stochastic learning [82]. An important issue concerning supervised learning is the problem of error convergence, i.e. the minimization of error between the desired and computed unit values. The aim is to determine weight values used for manipulating the inputs which minimizes the error. One well-known method, which is common to train the network on applying weights, is supply of weights as supplement to hidden neurons. The weights may be either constant or variable and in our model we supply constant weights. The two constant weight values that have been used are i) the age risk threshold and ii) the finance risk threshold which are discussed more elaborately in chapter 6.
3.8. PROCEDURE SEMANTICS OF THE PROPOSED COMPUTATIONAL MODEL

To the proposed model the procedural semantics has to be incorporated so the application scenario of the fusion model can be understood more precisely and hence for the various stages of the system design explained the working algorithm is drafted as below.

Procedure for Stage: 1

Rule Induction Algorithm

Input $D_t$ // Heterogeneous Crisp Database

Output $D_c$ // Homogeneous Crisp Database

Algorithm RI

// Induct rule over $D_t$;

Segregate $D_t$;

Get one or a collection of pruned crisp homogeneous database $D_c$;

As per procedure 3.8. any voluminous database is homogeneously classified so that the target of data search and analysis is narrowed down. In this modern information era the capacity of storage whether in the main memory or supplementary is of no matter but the very crucial concept of data mining, which is to assess and analyse the data serves to be the edge cutting factor in the applicatory aspects. This bottleneck event has been minimised by clustering...
databases of homogeneous kind by inducting supervised rule on the
unpruned crisp database. The result is the pruned crisp database Dc.

Procedure for Stage: 2

Fuzzy Transformation Algorithm

Input Dc // Homogeneous Crisp Database
Output D // Fuzzy Database

//Triangular membership functions are used to transform Dc to D;

Since this stage tries to develop a fuzzy associated data mining
methodology of illustrating the knowledge in the pertaining application
which mostly deals with quantitative values building the methodology
the fuzzy transformations using triangular membership functions are
used to transpose the pruned crisp database Dc into a fuzzy database D.

Procedure for Stage: 3

(i) Apriori Algorithm

Input: L_{i-1} // Large itemsets of size i-1
Output: C_i // Candidates of size i

Apriori Algorithm

C_i = 0;

for each I \in L_{i-1} do
  
  for each j \neq I \in L_{i-1} do
    
    if 1-2 of the elements in I and J are equal then
(ii) Adaptive - Fuzzy - Apriori (AFA) – Tree Algorithm on Samples

Input: I // Itemsets

D // Transformed Fuzzy Database

S // Support

Output: L // Potential itemsets

Algorithm AFA

K= 0; // k is used as the scan number

L = Ø;

C₁ = I; // Cardinality of items in the fuzzy database which are the initial candidates and are set to be the items

repeat

k=k+1;

Lₖ=Ø;

for each I, ∈ Cₖ do

Cᵢ=0; // Initial counts for all itemsets are 0

for each tᵢ ∈ D do

for each Iᵢ ∈ Cₖ do

    if Iᵢ ∈ tᵢ then

        Cᵢ = Cᵢ + 1;

for each Iᵢ ∈ Cₖ do
if \( C_i \geq (s \times D) \) do

\[
L_k = L_k \cup I_i;
\]

\[
L = L \cup L_k;
\]

\[
C_{k+1} = \text{Apriori} (L_k)
\]

until \( C_{k+1} = \emptyset \);

Hence to summarize on to the rule inducted sampling theory based AFA-Tree search algorithm the tested rules on the database improve the architecture of the database and the use of fuzzy transforms not only explains the far continuum of the real world values but tries to minimizes the number of iterations run by the general Apriori procedure. The rule of support is now framed to get the knowledge study of the dominant samples obtained by the procedure.

**Procedure for Stage:4**

**Multi layer Artificial Neural Network Rule Based Algorithm**

Input: \( D \) // Training data

\( N \) // Initial neural network

Output: \( R \) // Derived rules

Rule Based Algorithm

// generate rules that describe the output values in terms of the hidden activation values;
generate rules that describe hidden output values in terms of inputs;

combine the rules for validation.

The potential samples derived applying AFA by fixing up threshold for confidence is now trained using the single layer supervised ANN by rule based algorithm and now the ANN is ready for testing. The results are summarized for discussion.

3.9. EXPERIMENT

This hybrid model is now applied on the problem case which studies the customer side risk factors in investment in the Indian National Stock Market. As an example case the Indian perspective is taken and the money value is in terms of ‘Rupee’.

‘All of life is the management of risk, not its elimination’

-Walter Wriston, Former Chairman of Citicorp

Measuring and managing risk in an institution is a continuous process and not a one time activity. Thus in any sector, factors of risk have to be identified, measured, and mitigated. Simply put, risk, can be defined through Statistics as the ’Probability’ of loss which is percent honest in case of customer investment in stock market. In such a context this work tries finding an application paradigm using the
above described intelligent hybrid system by choosing the essential decision variables.

3.10. AIM OF THE EXPERIMENT

The aim of performing the experiment is

- To prove the applicatory efficiency of the proposed hybrid model.
- Proven results of minimal percentage of error.
- Efficient algorithm with minimum working iterations which reduces time complexity.
- Intelligent data assessment and analysis.
- Platform versatility and system compatibility.

3.11. SOFTWARE COMPATIBILITY OF THE HYBRID MODEL

The software run through the algorithm is developed in Sun Java which is a highly compatible language versatile over all platforms and is run on an Intel – Pentium IV improved configured machine.

3.12. SELECTION OF VARIABLES FOR DESIGN OF THE HYBRID MODEL

The most significant variables selected as data sets for approach of the design can be categorized into explicit variables and dependent
variables. To the problem approach the explicit variables such as the age of the investor, sex of the investor, amount of investment of the investor per transaction, term of investment, and type of investment play significant and very subjective role in decision making over the dependent variables like the historic values of the stock returns of the companies or sector where similar investments were made, the recurring income of the investor, sake of business and the sector profile. Looking further into the geometry of the decision variables, they are also quite subjective by nature which is the cause for getting into the synergy of the fuzzy domain in the design of the system.