Journal of Theoretical and Applied Information Technology
15th September 2016. Vol.91. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN ADAPTIVE PERSONNEL SELECTION MODEL FOR RECRUITMENT USING DOMAIN-DRIVEN DATA MINING

[1]MUHAMMAD AHMAD SHEHU, [2]FAISAL SAEED*

[1]Faculty of Computing. Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia

[2]Senior Lecturer, Faculty of Computing. Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia.

E-mail: [1]sdfax2014@gmail.com, [2]Faisalsaeed@utm.my

ABSTRACT

To support organizations in structuring personnel selection strategy for recruitment, various researches have been conducted using data mining approaches, and selection models containing selection rules were developed. Based on the methodology used, researches conducted were categorized as method – driven and domain–driven data mining approach of which domain–driven was discovered the preferred due to its model applicability in the real world. However, with the occasional changes in organization selection strategy, the models developed cannot adapt to these changes due to the static nature of the rules contained in the models. This research aims at developing an adaptive personnel selection model to support personnel selection for recruitment and adapt to the changes in personnel selection strategy. The framework used in developing the model involves Federal University Lokoja Nigeria recruitment dataset usage for extraction of selection rules to support personnel selection process using decision tree method of classification, generation of adaptive rules to handle the changes in personnel selection strategy using frequent and non-frequent pattern of data mining and domain expert's validation of each rule developed. The result of the implementation of the proposed model was ranked the highest after comparing it with selection models developed using four decision trees.

Keywords: *Human resource management, recruitment, personnel selection, data mining, method–driven data mining, domain-driven data mining.*

1. INTRODUCTION

The application of data mining in various human resource domains for decision support is increasing overtime due to the amount of datasets available in the domain that needs to be turned into meaningful information. Human resource is a managerial unit of an organization that concentrates on making managerial decision to influence the organization (personnel) performance through the organization efficiency or revenue growth [1]. The contributions of human resource are directly to the implementation of organization goals and objectives and it can be a unique source of sustained competitive advantage [2]. There are various sub domains that makes up the human resource domain but in this study the case sub domain is recruitment.

Recruitment is a human resource activity that deals with the selection of jobs/positions applicants into the organization. The selection of applicants is carried out based on a well – defined strategies structured by the organization. Recruiting of personnel in an organization affects the quality of employees, therefore selecting the right person for the right job is very important in human resource management. However, some organizations like a semiconductor industry [3][4], information technology industry [5][6] and so on, finds it tedious and complex to develop selection strategies for right personnel selection for recruitment based on their respective reason for recruitment. Due to this reason, researches like [3][4][5][6][7][8] have been carried out using data mining approach to develop suitable selection model (set of selection rules) as selection strategies from organizations personnel records to support the respective organizations in making decisions during selection of applicants for recruitment into the organizations.

However, due to the fact that selection strategies are changed due to various factors like change in organization work/job behavior, change in society

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

or organization laws and so on [9][10][11], the selection strategies (selection model or rules) developed through the solely dependent on the extraction of static rules from organization's personnel records can be used for personnel selection decision making until the need for changes in selection strategy of the respective organization arises. The selection model extracted will not be able to adapt to the need for change of selection strategy of the respective organization due to its static nature. This study aim at proposing a data mining framework for the development of personnel selection model that will be used for decision support and adapt to the change in selection strategies of organization. To demonstrate the validity of the proposed framework, an empirical study was carried out where the data mining framework was applied using a recruitment dataset of Federal University Lokoja, located in the capital of Kogi state of Nigeria. A selection model containing set of selection rules for supporting decision on recruiting academic applicants into the institution was developed and the procedure for the developed model to adapt to future changes in selection strategies was shown. A decision tree analysis, frequent and non-frequent pattern analysis was adopted in the framework.

2. RELATED WORK

Personnel selection during recruitment is carried out based on a well – structured selection strategy. In this paper selection model containing well – structured set of rules is represented as selection strategy. Personnel selection for recruitment depends on organization selection strategy, departments of organization selection strategy, and jobs of organization selection strategy [12]. The conventional method of recruitment involves activities like analyzing of candidates application form, telephone screening, self assessment and tests depending on the organization recruitment strategy and afterwards the candidates that have the profile that meets the organization selection rules are selected [6]. And with the help of information technology, as candidates are recruited the organization keeps records of these candidates in the Human Resource units. Some organizations keeps records of both applicant selected and rejected during recruitment exercise, while some only keep those that are selected.

At early 90s personnel selection approaches focuses on work and job analysis that are defined using specific tasks and duties based on their static behavior but according to [10][11] personnel selection practices are influenced by the changes in various area like organization, laws, society, marketing, technological advancement and so on. Therefore, personnel selection approaches using the idea of static work behavior is no longer suffice [13]. Moreover, Oswald (2000) found that the nature and analysis of work behavior also changes and can influence personnel selection. Some of the challenges in personnel selection include labor market shortages, technological developments, applicant perception of selection procedures and construct - driven approaches.

Due to the importance of personnel selection during recruitment, data mining specialists used data mining approach to develop several personnel selection model for recruitment to support several organization decisions in recruitment personnel. Wei-Shen and Chung-Chieng [8] used fuzzy data mining method on an existing transaction database of an organization to develop a tool to identify the relationship between applicant attributes and the organization job behavior. Chen–fu and Li–fei [3, 4] applied rough set theory data mining and the combination of classification and association methods of data mining respectively, guided by the information gathered from the human resource management experts of a semiconductor foundry company on the dataset collected from the company to develop a model for personnel selection during recruitment. Sivaram and Ramar [6] applied decision tree method of classification guided by the information gathered from the recruitment experts of an IT industry on the dataset collected from the industry to develop a model for personnel selection during recruitment.

Data mining is an information technology domain that uses combined domain knowledge of Statistics, Machine learning, and Artificial intelligence. It is sometimes referred to as Knowledge Discovery on Datasets. Data mining can be applied to various managerial domains like Customer Management [14], Financial Management [15], and Manufacturing Management [16]. The general approaches for solving data mining problems are Association, Classification, Clustering and Prediction approach [17] [18].

Recently, researchers discovered that the applicability approach of data mining to human resource domain and other managerial domain are categorized into method–driven data mining and domain–driven data mining approaches. Method–driven data mining approach is driven by the methods in data mining when discovering data mining solutions to domain problems, no other

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

knowledge is considered [19][20]. It is sometimes referred to as Data–driven data mining due to the solely dependency on data alone to generate solutions to various domain problems using data mining methods [21]. Due to the lack of involvement of the problem domain knowledge and validation of discovered solution during solution discovery in method–driven data mining approach, the solution discovered is said to have less applicability in the respective real world domain problems [19][20][21].

Domain–driven data mining is said to close the gap set by method–driven data mining approach [19][20][21]. The basic idea in domain–driven data mining is that it offers utility and relevance to data mining solutions to problems in various real world domains by carefully considering relevant problem domain–specific requirement and constraints during application of data mining processes [20]. Since domain – driven data mining approach involves the problem domain knowledge through the domain experts' intervention during solution discovery process, the approach is said to be preferred in solving human resource domain problems [20]. Based on this reason the approach used in this study is domain – driven data mining.

3. MATERIALS AND METHODS

3.1 Decision Tree Method of Classification

Decision tree to be constructed is a structured tree format of describing the classifications extracted from a dataset, where each of the internal node denote a test on an attribute, the branch describes the outcome of the test and the leaf represent the class labels. Classification rules can be generated from a decision tree [18]. There are various techniques of constructing a decision tree, but C4.5, Random Tree, REP Tree and CART decision trees were used in this paper. The selected decision tree methods are used due to the following reasons: Sivram and Ramar (2011) found that C4.5 decision tree method most accurately classified the case study dataset. And CART decision tree method is a powerful decision tree method due to its combined method (classification and regression) of classifying a dataset.

C4.5 is decision tree algorithm that uses gain ratio as splitting criterion to partition the dataset. REP Tree (Reduced – Error Pruned Tree) also known as the fast decision tree learner, construct decision tree or regression using information gain or variance and prunes the constructed tree using reduced–error pruning (with backfitting). While Random Tree construct a tree of *n* randomly chosen attributes at each node of the tree. During a Random Tree construction no pruning is carried out, however, it allows the opportunity for class probabilities (or target mean for the case of regression) based on hold – out set (backfitting). And CART (Classification and Regression Tree) method constructs a tree using both Regression and Classification data mining procedures. Tree pruning process is carried using statistical measures so as to remove anomalies created due to noisy data in the dataset used for constructing the tree.

Gain ratio is defined as in (1) and the attribute with maximum gain ratio is selected as the splitting attribute.

$$Gain\ ratio\ (A) = Gain\ (A)\ /\ Spilt\ info\ (A) \ldots\ldots \quad (1)$$

Split information for an attribute A of v attribute values is derived using (2)

$$Split\ info_A\ (D) = -\sum_{i=1}^{v}\frac{|Di|}{|D|} * \log\left(\frac{|Di|}{|D|}\right) \ldots\ldots\ldots \quad (2)$$

Where |Di| is the is the number of instances in the training set D with ith value for the attribute A and |D| is the total number of instances in the training set. The information gain for branching the dataset D based on the attribute A is obtained using (3).

$$Gain\ (A) = info\ (D) - info_A\ (D) \ldots\ldots\ldots\ldots\ldots \quad (3)$$

The expected information required for classifying a tuple in a dataset D is given as (4).

$$info\ (D) = -\sum_{i=1}^{m} Pi * log\ Pi \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

The probability Pi of an arbitrary tuple belonging to a class Ci of tuples in the dataset is measured as the ratio of number of tuples in the dataset that belongs to the class Ci to the total number of tuples in the dataset. The information needed after splitting the dataset based on attribute A of v attribute values is derived using (5).

$$info_A\ (D) = \sum_{j=1}^{v}\frac{|Dj|}{|D|} * info(D_j) \ldots\ldots\ldots\ldots\ldots \quad (5)$$

3.2 Dataset

Federal University Lokoja is an academic institution established alongside 8 other Universities in February, 2011. It is located at Lokoja, the capital city of Kogi State of Nigeria. Just like other academic institutions, it has the academic and non - academic units. The academic unit has two faculties: Faculty of Sciences with 6 departments and Faculty of Arts and Social Sciences with 5 departments. While the non-academic unit is made up of 20 units. The total

Journal of Theoretical and Applied Information Technology
15th September 2016. Vol.91. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

work force based on academic and non - academic units are 146 and 900 respectively (www.fulokoja.edu.ng).

Case study dataset was collected from Federal University Lokoja, and contains 266 academic applicants' profile that were selected (146) and rejected (120) during recruitment from the year 2011 to date. Each applicant profile contains attributes that describes the applicants, and these attributes are used in justifying the qualification for the position applied for by the applicant. Each applicant profile contains 17 attributes used to justify the selection of the applicants.

However, according to the problem domain expert, not all attributes in the applicants' profile are used to justify the qualification of the applicants, so therefore feature selection process of the data preprocessed stage was carried out based on the selection of the problem domain expert. Table 6 contains selected features (attributes) majorly used for selection justification, out of the 17 attributes of the applicants' profile.

Table 1 Applicants attributes used for selection justification

S/N0	Selected attributes
1	Post applied
2	Previous organization relationship
3	Academic qualifications obtained
4	Teaching and research experience
5	Publications
6	Honours and awards
7	Credential screening

Some of the attributes were in nominal form but were converted to ordinal (categorical) form manually. Some attributes were constructed from existing attributes like the department attribute were constructed from the post applied attribute, since each of the attribute value of the post applied has the attachment of the department the applicant applied into. Outliers (noisy data) were removed manually. Table 2 describes the preprocessed attributes obtained in this stage.

Table 2 Attributes and their respective initial data types

S/N0	Constructed attributes	Initial Data type
1	Post applied	Ordinal
2	Department	Ordinal
3	Academic qualification	Ordinal
4	Teaching and research experience	Continuous
5	No of Publications	Continuous
6	No of academic awards	Continuous
7	Any bond with other organization	Ordinal
8	Credentials screening status	Ordinal

3.3 Evaluation Methods

There are various methods of evaluation, but accuracy, error rate, precision and recall are the evaluation measurements adopted in this study.

The 10 fold cross-validation was used, such that the initial data are randomly partitioned into 10 mutually exclusive subsets or "folds," D1, D2,…, D10, each of approximately equal size. Training and testing is performed 10 times. In iteration i, partition Di is reserved as the test set, and the remaining partitions are collectively used to train the model. That is, in the first iteration, subsets D2,…. , D10 collectively serve as the training set to obtain a first model, which is tested on D1; the second iteration is trained on subsets D1, D3, …. , D10 and tested on D2, and so on. The evaluation measurements for accuracy (percentage of correctly classified set of tuples) is given in Equation 6, error rate (complement of accuracy) is given in Equation 7, precision (percentage of tuples correctly classified by the model as positive (selected) or negative (rejected) out of positive (selected) class or negative (rejected) class of tuples respectively) is given in Equation 8 and recall (percentage of correctly classified tuples by the model as positive or negative) is given in Equation 9.

Assuming in a dataset D containing n total records, TP represents number of tuples that are correctly classified as positive (selected) by a model M, TN represents number of tuples that are correctly classified as negative (rejected) by M, FP

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

represents the number of tuples that are incorrectly classified as positive (selected) and FN as number of tuples incorrectly classified as negative (rejected), then:

$$Accuracy\ \%\ (M) = ((TP + TN) / n) * 100 \ldots\ldots (6)$$

$$Error\ rate\ \%\ (M) = 100 - Accuracy\ \%\ (M) \ldots.. (7)$$

$$Precision\ \%\ (M) = (TP / (TP + TN)) * 100 \ldots. (8)$$

$$Recall\ \%\ (M) = (TP / (TP + FN)) * 100 \ldots\ldots.. (9)$$

Also in this study Kendall's significance test is used during final model evaluation and analysis. Kendall's significance test is a statistical measure of the agreement among several judges who are assessing a given set of *n* objects [22]. The objects in this study are evaluation measurements (accuracy, error rate, precision and recall), the judges are personnel selection models developed using this study's framework and the decision tree methods.

4. EXPERIMENTAL DESIGN

The experimental design of this study involves the application of the proposed framework to develop a personnel selection model to support recruitment decisions and to adapt to the changes in personnel selection strategy during recruitment. The development of the proposed model (adaptive personnel selection model) majorly depends on the processes carried out in phase 1 to 3 of the study framework. The adaptive nature of the proposed personnel selection model depends on its implementation described in phase 4 of the framework. As shown in Figure 1, the framework consists of four phases; preliminary phase, decision tree based personnel selection rules extraction phase, adaptive rules generation phase and Implementation of adaptive personnel selection model phase.

4.1 Phase1: Preliminary Phase

This phase consist of two preliminary stages: Knowledge acquisition and data preprocessing stages. At the end of this phase all the necessary tools and information needed to carry out the experimentation must have been collected.

4.2 Phase 2: Decision Tree Base Personnel Selection Rules Extraction

This phase consists of 5 stages; construction of four decision tree, evaluation of constructed decision trees, extraction of rules from most accurate decision tree, expert validation of extracted rules and evaluation of extracted rules stages. The rules extracted at the end of this phase are personnel selection rules used by the Institution from past to present period of recruitment. This phase uses the proposed method of Sivram and Ramar, (2010) to extract selection rules from recruitment dataset, however, in this study, after validation of the extracted rules the validated model (set of validated rules) was evaluated. This phase is carried out based on the idea that, it is recommended to use personnel selection rules extracted from past recruitment dataset to support current personnel selection decisions during recruitment [4, 6].

Construction of four decision tree involves application of C4.5, RandomTree, REPTree and CART decision tree on the preprocessed dataset and then after evaluation of the four decision tree models in the evaluation stage, the most accurate decision tree was used for further experimentation.

Personnel selection rules are extracted from the most accurate decision tree algorithm. And these extracted rules are reviewed by experts in academic personnel recruitment and the most appropriate rules that solve issues in academic personnel selection during recruitment are selected. The collection of the validated rules is a selection model to be used for personnel selection during recruitment in the case study.

In the evaluation of validated model (set of validated rules) stage, the validated model is evaluated based on the evaluation measurements of accuracy, error rate, precision and recall, then compared with any case study evaluation threshold. The validated model can be used to support personnel selection decision making if the model (set of validated rules) meets the evaluation threshold given by the case study problem domain expert.

4.3 Phase 3: Adaptive (Dynamic) Rules Generation

This phase of the framework consists of four stages; determination of frequent attributes, determination of non–frequent attributes, adaptive rules derivation and expert's validation of adaptive rules derived. However, before the commencement of the first stage, the preprocessed dataset is segmented based on the selected and rejected class labels afterwards the segment containing selected tuples is collected. The collected segment is considered based on the fact that the applicants'

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attributes (profile) are compared with rules to be selected not rejected that is, if applicants meet the rules to be selected the applicant is selected if not applicant will be rejected.

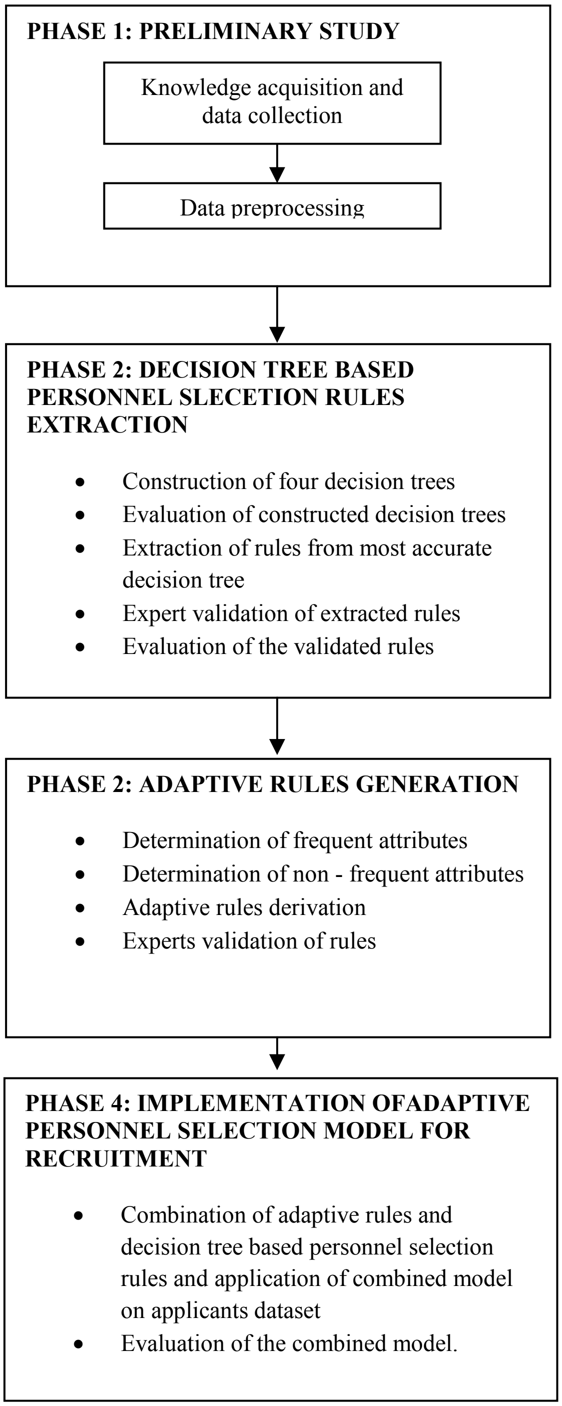

Figure 1 Framework for adaptive personnel selection model development

4.3 Phase 3: Adaptive (Dynamic) Rules Generation

This phase of the framework consists of four stages; determination of frequent attributes, determination of non–frequent attributes, adaptive rules derivation and expert's validation of adaptive rules derived. However, before the commencement of the first stage, the preprocessed dataset is segmented based on the selected and rejected class labels afterwards the segment containing selected tuples is collected. The collected segment is considered based on the fact that the applicants' attributes (profile) are compared with rules to be selected not rejected that is, if applicants meet the rules to be selected the applicant is selected if not applicant will be rejected.

The scope of the study framework in developing selection rules is based on job selection strategy, therefore the collected selected segment is further divided into selected dataset segments based on the positions or jobs applied by the applicants (that is the number of selected dataset segments is equal to the number of position applied). This implies that the adaptive rules to be generated are with respect to the positions (that is adaptive rule generated in a selected dataset segment is the adaptive rule generated for position in that particular selected dataset segment). This phase is recursively carried out with respect to the number of selected dataset segments.

For each selected dataset segment, all the frequent attributes are separated from non – frequent collection of attributes. The idea behind this process is to identify part of the selection rules attributes that does not change overtime based on the dataset collected (frequent attributes) and part of the selection rules attributes that changes overtime based on the dataset collected (non – frequent attributes).

4.3.1 Determination of frequent and non - frequent attributes

An attribute is said to be frequent if its value is frequent. Therefore to identify the attribute or collection of attributes that are frequent in each selected dataset segment, respective attributes values frequency are measured based on a frequency threshold (100%, because if less than 100% then there is a certain period of personnel selection for recruitment where the respective attribute had a change of attribute values which implies that the attribute had not been frequent).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Attribute value frequency estimation is given in (10).

If dataset *D* contains *n* tuples, for an attribute *a* with attribute value *v*, the frequency *f* of *v* in *D* is given as:

$$f(v)\ \% = ((\text{number of } v \text{ in } D) / n) * 100 \quad (10)$$

If an attribute has more than one attribute value, then each of the attribute values frequency is measured. If any of the attribute values satisfies the frequency condition, then the respective attribute is identified as a frequent attribute.

All the attributes that are not selected during determination of frequent attributes stage are identified as non–frequent attributes.

4.3.2 Adaptive rules derivation and expert's validation of adaptive rules

The frequent attributes and non–frequent attributes selected in their respective stages are to be used in generating adaptive rules in this stage. The frequent and non–frequent attributes extracted from a particular selected dataset segment are to be used in generating adaptive rule for that particular selected dataset segment. As such, the adaptive rule generated from a selected dataset segment is the adaptive rule for the job or position in that particular selected dataset segment. Just as described in Table 3, the "IF - THEN" rule format is used in constructing the adaptive rules, where the attributes identified to be frequent are tested with their respective attribute values that made them frequent and the attributes identified as non – frequent are tested with values to be inputted by the case study based on the change in selection strategy.

Table 3 Adaptive (dynamic) rule sample

RULE1	**IF** (Att1 = Y) **AND** (Att2 = Y) **AND** (Att3 = Institution input) **AND** (Att4 = Institution input) **AND** (Position applied = Position2) **THEN Selected**

In Table 3, Att1, Att2 and Position applied attributes are identified as frequent attributes and Y, Y and Position2 are their respective attribute values that made them frequent, while Att3 and Att4 are identified to be non–frequent attributes, therefore tested with values to be inputted by the case study. For each adaptive rule generated in this stage, the constant part of the rule represents the criteria that must be owned by all applicants or personnel applying for the respective job/position due to the fact that it is a common (frequent) criterion or criteria for all the personnel in that job/position and throughout the period of personnel selection for recruitment, the common (frequent) criterion (attribute) or criteria (attributes) did not change. And the input part of the rule represents the selection criteria that changes overtime.

The adaptive rules derived were reviewed by the domain experts to justify the meaningfulness of the rules before implementing them. If for any reasons, rules generated are far from common practice, then further investigation will be carried out to validate the rules.

4.4 Phase 4: Implementation of Adaptive Personnel Selection Model for Recruitment

This phase involve the combination of the decision tree based personnel selection rules extracted and validated and adaptive (dynamic) rules generated and validated.

4.4.1 Combination of adaptive rules and decision tree based personnel selection rules

This stage describes how the adaptive personnel selection model is implemented for recruitment. This stage is continuous for every change in selection rules that occurs. And will require inputs from the organization based on the changes. Figure 2 describes the flow of this stage (how the model that will adapt to changes in selection strategy is developed and used).

If any change in personnel selection rules for any job or position should occur during recruitment, the adaptive rules for the positions with change in personnel selection rules are selected then the required institution inputs based on the changes in selection rules for the respective positions are entered in the adaptive rules selected to become new selection rule or rules, afterwards the rule or rules are combined with the decision tree based personnel selection rules to become the new personnel selection model for recruitment (combined personnel selection model). This implies that for every implementation of the adaptive personnel selection model due to change in personnel selection strategy, a new combined personnel selection model is achieved due to the new selection rules generated from the adaptive rules and combined with the current combined personnel selection model.

Journal of Theoretical and Applied Information Technology
15th September 2016. Vol.91. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

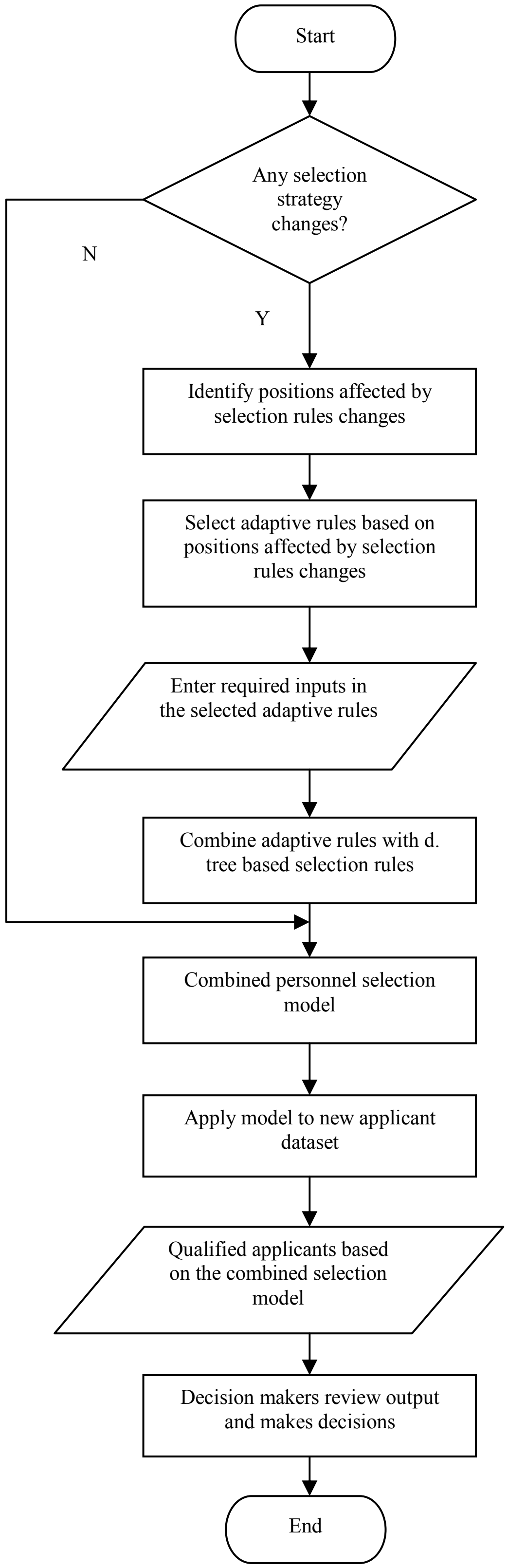

Figure 2: Flow for combining decision tree rules and adaptive rules

Assuming a decision tree based personnel selection rules representing a personnel selection model used for decision support is given in Table 4 and the adaptive rules generated in the adaptive rules generation phase is given in Table 5.

Table 4 Sample of decision tree based selection rules

RULE1	**IF** ((Att1 = Y) **AND** (Att2 = Y) AND (Att3 = L) AND (Position applied = Position2)) **THEN Selected**
RULE2	**IF** (Att1 = N) **THEN Rejected**
RULE3	**IF** ((Att1 = Y) **AND** (Att2 = N) **AND** (Position applied = Position3)) **THEN Rejected**

Table 5 Sample adaptive rules

RULE1	**IF** (Att1 = Y) **AND** (Att2 = Y) **AND** (Att3 = Institution input) **AND** (Att4 = Institution input) **AND** (Position applied = Position2) **THEN Selected**
RULE2	**IF** ((Att1 = Y) **AND** (Att2 = Y) **AND** (Att3 = H) **AND** (Att4 = Institution Input) **AND** (Position applied = Position3)) **THEN Selected**
RULE3	**IF** (Att1 = Y) **AND** (Att2 = Y) **AND** (Att3 = Institution input) **AND** (Att4 = Institution input) **AND** (Position applied = Position2) **THEN Selected**

Assuming there is a new rule for selecting personnel into position3, before combination, the adaptive rule where Position applied is "position3" is selected (Rule2 of sample adaptive rules (Table5)). The input needed in the adaptive rule selected is value for Att4 (assuming <5), then the new rule for selecting personnel into position3 generated from Rule2 is given in Table 6. Combining the new rule for position3 with the decision tree based personnel selection rules gives a selection model given in Table 7.

Table 6 New position 3 selection rule

RULE1	**IF** ((Att1 = Y) **AND** (Att2 = Y) **AND** (Att3 = H) **AND** (Att4 < 5) **AND** (Position applied = Position3)) **THEN Selected**

Table 7 Combined selection model

RULE1	**IF** ((Att1 = Y) **AND** (Att2 = Y) AND (Att3 = L) AND (Position applied = Position2)) **THEN Selected**
RULE2	**IF** (Att1 = N) **THEN Rejected**
RULE3	**IF** ((Att1 = Y) **AND** (Att2 = N) **AND** (Position applied = Position3)) **THEN Rejected**
RULE4	**IF** ((Att1 = Y) **AND** (Att2 = Y) **AND** (Att3 = H) **AND** (Att4 < 5) **AND** (Position applied = Position3)) **THEN Selected**

It should be noted that if the selection rules or strategy are not changed during recruitment, then the adaptive rules are ignored, instead the personnel selection model (set of rules) that was previously used during the last recruitment is used. For instance, if Table 7 is used as the model for personnel selection during recruitment and the next recruitment process do not require any changes in personnel selection rules then Table 7 will continuously be used until changes in personnel selection rules occurs.

The implementation of the adaptive personnel selection model was practically carried out to validate the usefulness of the model.

5. RESULTS AND DISCUSSIONS

The problem identified in the case study is the task of selecting qualified academic applicants by the human resource unit. To select the right person for the right job is a complex task due to the amount and inconsistency of the applicants' profiles to be examined, and generating a well – defined set of selection rules to be used. To solve this problem, this study developed a selection model for the selection of qualified academic applicants.

Based on the framework of this study, the necessary idea in selection of applicants was acquired through frequent interview with the case study recruitment expert. However, the knowledge acquisition step was a continuous process until the final model was achieved.

5.1 Decision Tree Based Personnel Selection Rules Extraction Phase

Four decision trees (C4.5, Random Tree, REP Tree and CART) were constructed using the preprocessed dataset. WEKA 3.7.13 data mining tool was used during construction of the decision tree algorithms and evaluation of each algorithm. Table 8 shows the evaluation measures for each decision tree model constructed using the preprocessed dataset. C4.5 decision tree method was discovered to construct the most accurate decision tree model, therefore 26 selection rules were extracted from the C4.5 decision tree model. Table 9 contains some of the selection rules extracted from the C4.5 decision tree.

Table 8 Not validated Decision tree models evaluation estimates

Decision tree model	Accuracy (%)	Error rate (%)	Precision (w. Avrg) %	Recall (w. Avrg) %
C4.5	98.4%	1.6%	98.5%	98.5%
RandomTree	92.9%	7.1%	93.1%	92.9%
REPTree	93.6%	6.4%	93.7%	93.6%
CART	95.1%	4.9%	95.4%	95.1%

Table 9 C4.5 decision tree based rules

RULE1	**IF** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (Credentials screening status = PASS) **AND** (Any bond with other org = NO) **THEN SELECTED**
RULE2	**IF** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (Credentials screening status = PASS) **AND** (Any bond with other org = YES) **THEN REJECTED**
RULE3	**IF** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (Credentials screening status = FAIL) **THEN REJECTED**

The 26 selection rules extracted from the C4.5 decision tree model were reviewed by the case study domain expert (The head of department of Computer Science: Involved in most of the recruitment of academic staff in the institution) and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

19 rules were validated to be applicable for selecting academic personnel or applicants. However, 2 of the validated selection rules were observed not to be well structured due to the fact that no position specified in the 2 rules, therefore the 2 rules were also removed, as such the validated selection rules became 17. And when the validated C4.5 based selection model (validated set of selection rules) was evaluated, the accuracy estimate was 88.7% as in Table 11, and based on the domain expert, the personnel selection model accuracy should not be less than 90%, this implies that the selection model extracted has accuracy below the accuracy threshold given and as such needs to be improved. Table 10 shows some of the validated rules.

Table 10 Validated C4.5 decision tree based rules

RULE2	**IF** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (Credentials screening status = PASS) **AND** (Any bond with other org = YES) **THEN REJECTED**
RULE3	**IF** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (Credentials screening status = FAIL) **THEN REJECTED**
RULE5	**IF (**Academic qualifications = DEGREE, M DEGREE**) AND (**Credentials screening status = PASS) **AND** (Any bond with other org = NO) **AND (**Post applied = ASSISTANT LECTURER**) THEN SELECTED**
RULE6	**IF (**Academic qualifications = DEGREE, M DEGREE**) AND (**Credentials screening status = PASS) **AND** (Any bond with other org = NO) **AND (**Post applied = LECTURER2) **AND (**No OF PUBLICATIONS <= 1) **THEN REJECTED**
RULE7	**IF (**Academic qualifications = DEGREE, M DEGREE**) AND (**Credentials screening status = PASS) **AND** (Any bond with other org = NO) **AND (**Post applied = LECTURER2) **AND (**No OF PUBLICATIONS > 1) **THEN SELECTED**

However, since this study framework gives the opportunity to use adaptive set of rules to restructure a personnel selection model by adding new selection rules generated by the adaptive rules, then the evaluation estimation for accuracy was improved using the adaptive rules generated. For further analysis, selection rules were extracted, validated and evaluated from the rest 3 decision tree algorithms (RandomTree, REPTree and CART). The evaluation estimates for the validated selection models (set of selection rules selected by the expert) extracted from each decision tree algorithm are given in Table 11.

Table 11 Validated decision tree based selection model evaluation measures

Decision tree model	Accuracy (%)	Error rate (%)	Precision (w. Avrg) (%)	Recall (w. Avrg) (%)
C4.5	88.7%	11.3%	89.5%	89.4%
Random Tree	82.7%	17.3%	82.9%	82.8%
REPTree	78.9%	21.1%	79%	79%
CART	71.7%	28.3%	71.8%	72.4%

5.2 Adaptive Rules Generation Phase

The preprocessed dataset contains 266 applicants' profiles, dividing the preprocessed dataset based on the process described in the framework before the first stage of the adaptive rules generation phase results to the breakdown in Table 12. Each stage of the adaptive rules generation phase is recursively carried out on each selected dataset segments.

Table 12 Preliminary process results for adaptive rules generation phase

Dividing 266 applicants' data based on selected and rejected applicants		
Dataset segment 1	Selected applicants	146
Dataset segment 2	Rejected applicants	120
Dividing Selected applicants data based on 7 position applied		
Selected dataset segment 1	Graduate Assistant (GA)	44
Selected dataset segment 2	Assistant Lecturer (AL)	56
Selected dataset segment 3	Lecturer2 (L2)	5
Selected dataset segment 4	Lecturer1 (L1)	9
Selected dataset segment 5	Senior Lecturer (SL)	16
Selected dataset segment 6	Reader/Associate Prof (R/AP).	3
Selected dataset segment 7	Professor (P)	13

Table 13 contains the number of frequent and non- frequent attributes for each selected dataset

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

segment. Since we have 7 selected dataset segments then 7 adaptive rules were generated. Table 14 contains some of the adaptive rules generated.

Table 13 Number of frequent and non-frequent attributes in all selected dataset segments

Selected dataset segments	Number of Frequent Attributes	Number of Non - frequent Attributes
Selected dataset segment 1	6 out of 8	2 out of 8
Selected dataset segment 2	4 out of 8	4 out of 8
Selected dataset segment 3	4 out of 8	4 out of 8
Selected dataset segment 4	3 out of 8	5 out of 8
Selected dataset segment 5	4 out of 8	4 out of 8
Selected dataset segment 6	4 out of 8	4 out of 8
Selected dataset segment 7	4 out of 8	4 out of 8

Table 14 Adaptive rules for selected dataset segment 1, 2

Selected dataset segments	Adaptive rules
1	**IF** (Post applied = Graduate Assistant) **AND** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (No of publications = 0) **AND** (Teaching and research experience (Yes) = 0) **AND** (Credential screening status = PASS) **AND** (Any bond with other organization = NO) **AND** (Department = "Institution input") **AND** (No of academic awards = "Institution input") **THEN SELECT**
2	**IF** (Post applied = ASSISTANT LECTURER) **AND** (Academic qualifications = Degree, M Degree) **AND** (Credential screening status = PASS) **AND** (Any bond with other organization = NO) **AND** (No of publications = "Institution input") **AND** (Teaching and research experience = "Institution input") **AND** (Department = "Institution input") **AND** (No of academic awards = "Institution input") **THEN SELECT**

Each adaptive rule generated from its respective selected dataset segment as in Table 14 represents the adaptive rule for selecting applicants applying for the position in the respective dataset segment (that is adaptive rule generated in selected dataset segment1 represent adaptive rule for selecting applicants applying for Graduate Assistant position and so on)

And all the 7 adaptive rules generated were validated to be applicable depending on the input used in the rules.

5.3 Combination of the Adaptive Rules and C4.5 Decision Tree Base Personnel Selection Rules

To restructure the personnel selection model extracted in the decision tree based personnel selection rules extraction phase so as to improve the accuracy of the model, some of the invalidated rules like the 2 selection rules that were removed due to their unstructured nature (no position specified in the rules) were restructured using the adaptive rules generated by identifying which positions the rules (the 2 unstructured selection rules) are applicable using the directions of the domain expert (4 positions were identified to be in connection with the 2 unstructured selection rules), then select the adaptive rules for the positions identified and then enter the inputs of the adaptive rules using the attribute values of the 2 unstructured selection rules. Then combine the restructured rules with the validated C4.5 decision tree based rules to become a combined personnel selection model. Table 15 describes some of the rules contained in the combined personnel selection model where Rule1 to Rule3 are the C4.5 decision tree based rules and Rule18 to Rule21 are restructured rules derived from the adaptive rules. The evaluation estimates for the combined personnel selection model is given in Table 16.

It was observed that the accuracy threshold (90% and above) was met after evaluating the combine personnel selection model, therefore making the combine personnel selection model valid to be use for supporting decisions during selecting of academic applicants into the institution.

To verify the significance of this study, the personnel selection model (combined personnel selection model) evaluation estimations was compared with the evaluation estimations (Table 8 and Table 11) of the personnel selection models derived from the 4 decision tree methods (C4.5, RandomTree, REPTree and CART) during the

Journal of Theoretical and Applied Information Technology
15th September 2016. Vol.91. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Extraction of decision tree based personnel selection model phase of this study's framework. Table 16 shows the evaluation estimations of all the personnel selection models developed in this study.

Table 15 Combined personnel selection model

RULE1	**IF** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (Credentials screening status = PASS) **AND** (Any bond with other org = YES) **THEN REJECTED**
RULE2	**IF** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (Credentials screening status = FAIL) **THEN REJECTED**
RULE3	**IF (**Academic qualifications = DEGREE, M DEGREE**) AND (**Credentials screening status = PASS) **AND** (Any bond with other org = NO**) AND (**Post applied = ASSISTANT LECTURER**) THEN SELECTED**
RULE18	**IF** (Post applied = Graduate Assistant) **AND** (Academic qualifications = SECOND CLASS UPPER DEGREE OR ABOVE) **AND** (No of publications = 0) **AND** (Teaching and research experience (Yrs) = 0) **AND** (Credential screening status = PASS) **AND** (Any bond with other organization = NO) **AND** (Department = Any department) **AND** (No of academic awards >= 0) **THEN SELECT**
RULE19	**IF** (Post applied = SENIOR LECTURER.) **AND** (Academic qualifications = DEGREE, M DEGREE, PH DEGREE) **AND** (No of publications >= 5) **AND** (Teaching and research experience > 10) **AND** (Credential screening status = PASS) **AND** (Any bond with other organization = NO) **AND** (Department = Any department) **AND** (No of academic awards > 0) **THEN SELECT**
RULE20	**IF** (Post applied = READER/ASSOCIATE PROF.) **AND** (Academic qualifications = DEGREE, M DEGREE, PH DEGREE) **AND** (No of publications >= 12) **AND** (Teaching and research experience > 10) **AND** (Credential screening status = PASS) **AND** (Any bond with other organization = NO) **AND** (Department = Any department) **AND** (No of academic awards > 0) **THEN SELECT**
RULE21	**IF** (Post applied = PROFESSOR) **AND** (Academic qualifications = DEGREE, M DEGREE, PH DEGREE) **AND** (No of publications >= 20) **AND** (Teaching and research experience > 10) **AND** (Credential screening status = PASS) **AND** (Any bond with other organization = NO) **AND** (Department = Any department) **AND** (No of academic awards > 0) **THEN SELECT**

Table 16 Experiments results (selection models) evaluation measures

	Acc (%)	**E. rate (%)**	**Prec (w. Avrg) %**	**Recall (w. Avrg) %**
Combined personnel selection model	99.6%	0.4%	99.6%	99.7%
Not validated C4.5 personnel selection model	98.4%	1.6%	98.5%	98.5%
Validated C4.5 personnel selection model	88.7%	11.3%	89.5%	89.4%
Not validated RandomTree personnel selection model	92.9%	7.1%	93.1%	92.9%
Validated RandomTree personnel selection model	82.7%	17.3%	82.9%	82.8%
Not validated REPTree personnel selection model	93.6%	6.4%	93.7%	93.6%
Validated REPTree personnel selection model	78.9%	21.1%	79%	79%
Not validated CART personnel selection model	95.1%	4.9%	95.4%	95.1%
Validated CART personnel selection model	71.7%	28.3%	71.8%	72.4%

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

It was observed that the Combined personnel selection model (study model), and all the Not validated (because the rules contained in the models have not been validated) personnel selection models in Table 16 satisfies the case study personnel selection model accuracy threshold (90% and above), however, the Combined personnel selection model has more accuracy and other evaluation estimations than all the models that satisfies the accuracy threshold. However, human resource data mining demands that results of any application of data mining method on any human resource problem domain be validated for applicability to the respective problem domain (domain – driven data mining approach) [20], therefore based on the applicability of the given models in Table 16 in the domain of academics, the Not validated set of personnel selection models are lacking due to the fact that the set of rules contained in the models are not validated. Although, when the Not validated set of personnel selection models were validated their respective accuracy reduced below the case study accuracy threshold. This implies that the combined personnel selection model contains more accurate and applicable rules in the domain of academic than the rest compared models due to the fact that the rules contained in the model are validated and the model accuracy satisfies the case study accuracy threshold.

Based on the case study dataset and the evaluations in Table 16, it is concluded that this study framework produces a more accurate and validated personnel selection model for recruitment due to its adaptive nature compared to the used decision tree classification methods due to their static nature.

Using the Kendall's significance ranking shown in Table 17, this study model is ranked as the highest based on 4 different properties (accuracy, error rate, precision and recall).

The models are listed in rank descending order as: Combined personnel selection model > Not validated C4.5 personnel selection model > Not validated CART personnel selection model > Not validated REPTree personnel selection model > Not validated RandomTree personnel selection model > Validated C4.5 personnel selection model > Validated RandomTree personnel selection model > Validated REPTree personnel selection model > Validated CART personnel selection model. IBM SPSS Statistics 22 was the statistical tool used in achieving the Kendall's rankings in Table 17.

Table 17 Kendall's significance rankings for each model

	Mean Rank
Combined_model	7.00
NV_C4.5_model	6.50
V_C4.5_model	4.50
NV_RandomTree_model	5.00
V_RandomTree_model	4.00
NV_REPTree_model	5.50
V_REPTree_model	3.50
NV_CART_model	6.00
V_CART_model	3.00

6. CONCLUSION

This study contribution to data mining application to discover solution to problem domain in management is focused on recruitment data mining, where selection model is developed to support decisions during selecting applicants for recruitment.

Various selection models have been developed in previous research contributions, but due to the static nature of the selection rules in the model developed, if selection rules should be changed the selection model developed will not be able to adapt to the changes. However, this study proposes a domain-driven data mining framework to be used for developing selection model that can be used for decision support and can adapt to the change in selection rules when it occurs.

The study framework was applied to a case study (Federal University Lokoja) and a personnel selection model was developed. It was observed during the study comparison with 4 decision tree methods in developing personnel selection model using the case study dataset, that the proposed model gave a more accurate and validated personnel selection model and when all methods model were ranked using the Kendall's significance test, this proposed model was ranked the highest.

Acknowledgments. This work is supported by the Ministry of Higher Education (MOHE) and Research Management Centre (RMC) at the Universiti Teknologi Malaysia (UTM) under the Fundamental Research Grant Scheme (FRGS) Category (VOT R.J130000.7828.4F741).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES

[1] B. Becker, and B. Gerhart. The Impact of Human Resource Management on Organizational Performance: Progress and Prospects. *The Academy of Management Journal*. Vol. 39, 1996, No. 4, pp 779 – 801.
[2] L. Baird and I. Meshoulam. Managing two fits of strategic human resource management. *Academy of Management review*, 13(1), 1988, 116-128.
[3] C. Chen-Fu and C. Li-Fei. Data mining to improve personnel selection and enhance human capital. *Expert systems with applications*. 34, 2008, 280 – 290.
[4] C. Chen-Fu and C. Li-Fei. Using rough set theory to recruit and retain High – potential talents for semi conductor manufacturing. *IEEE Transactions on semi conductor manufacturing*. Vol 20, 2007, No 4.
[5] N. Sivaram and K. Ramar. Applicability of Clustering and Classification algorithms for Recruitment data mining. International *Journal of Computer applications*. vol. 4, 2010, No 5 0975 – 8887.
[6] N. Sivaram and K. Ramar. Knowledge Engineering to aid the recruitment process of an Industry by identifying superior selection criteria. *ICTACT Journal on soft computing.* 2011, ISSN: 2229 – 6956 (online).
[7] H. Min and A. Emam. Developing The Profile of Truck drivers For Their Successful Recruitment and Retention: A data mining approach. *International Journal of Physical Distribution and Logistics Management*. 33, 1/2; ABI/INFORM Global, 2003, pg. 149.
[8] T. Wei-Shen and H. Chung-Chian. A realistic Personnel Selection ToolBased on Fuzzy Data mining Method. *JCSI. Proceeding*. 2006, ISSN. 1951 – 6851.
[9] F. Oswald. Personnel selection: Looking towards the future remembering the past. *Annual Review of Psychology*. 51, 2000, 631-664.
[10] I. Robertson and M. Smith. Personnel Selection. *Journal of Occupational and Organizational Psychology*. 74, 2001, 441-472.
[11] C. Borman and M. Hanson. Personnel Selection. *Annual Review of Psychology*. Vol 48, 1997, 299-337
[12] C. Lewis. *Employee selection*. UK: Nelson Thrones Ltd., London, 1987.
[13] F. Lievens, K. van Dam and N. Anderson. Recent trends and challenges in personnel selection. *Personnel Review*, 31(5), 2002, 580-601.
[14] E. Ngai, L. Xiu and D. Chau. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36, 2009, 2592–2602.
[15] A. Choudhary, J. Harding and M. Tiwari. Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20, 2008, 501–521.
[16] E. Ngai, Y. Hu, Y. Wong, Y. Chen and X. Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50, 2011, 559–569
[17] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smith and R. Uthurusamy. Advances in Knowledge Discovery and Data Mining, 1996, 1-34. *AAAI.*
[18] H. Jiawei and M. Kamber. *Data mining: concepts and techniques*. San Francisco, CA, itd: Morgan Kaufmann, 2001, 5.
[19] A. Adeyemi and M. Amir. Domain Driven Data mining – Application to Business. *International Journal of Computer Science Issues*. Vol. 7, Issue 4, 2010, No 2.
[20] F. Stefan and P. Franca. Domain - driven data mining in human resource management: A review of current research. *Expert system with applications*. 40, 2013, 2410-2420.
[21] M. Kumari. Data Driven Data Mining to Domain Driven Data Mining. *Global Journal of Computer Science and Technology.* Vol.11, issue 23, 2011, Online ISSN: 0975 – 4172.
[22] P. Legendre. Species Associations: The Kendall Coefficient of Concordance Revisited. *Journal of Agricultural, Biological, and Enviromental Statistics*. Vol. 10, 2, 2005, 226-245.