

Combining Fuzzy Rough Set with Salient Features for HRM Classification

Asia L. Jabar

Computer Science, College of Science,
University of Sulaimani
Sulaimani, Kurdistan
asia.jabar@yahoo.com

Tarik A. Rashid

Software Engineering, College of Engineering,
Salahadin University-Erbil
Hawler, Kurdistan
tarik.rashid@su.edu.krd

Abstract—in today's economic transformation setting round the globe, there has been a growing interest in Human Resources Management (HRM) of corporations and their consequence on revenues of these corporations. Yet, there are some challenges and issues in deciding about the best people with talents and recommending them for rising in financial gain or promotion based on some features which are vital for the interests of the corporations.

This paper presents a solution for Human Resource Talent Management (HRTM) problem via using data mining techniques. In this research work, effective feature selection methods are used, then, the classification task is conducted via Fuzzy Rough Nearest neighbors, Decision Tree and Naïve Bayes. Basically, the information gained by using combining filter feature selection techniques then using Fuzzy Rough Set theory and depending on the results, Fuzzy Rough Nearest Neighbors classifier has the highest classification accuracy rate (which was 98.1174%) among others.

Keywords—Fuzzy Rough Nearest Neighbors, Decision Tree, Naïve Bayes, Fuzzy Rough Set Theory, Feature Selection, HRTM.

I. INTRODUCTION

In any organization, HRM is considered to be the most central strength for a corporation to be competitive and economical. It becomes conceivable for an employee to leave his or her job due to the liberalization on the labor market, and certainly, the employee allocation will be more effectual and resourceful when the employees of a department change habitually. In addition, the assurance and morale of the companies will be swayed when having excess employees leave their jobs; this is called Snowball Effect which means an employee's leaving will persuade his coworkers to leave one by one. This kind of trend unquestionably would have a great weight upon the operation of the corporation. Improvement and invention of an artifact could be reproduced, but capable teamwork and workforces cannot be replicated. The competitive advantage of a corporation can be weakened through the loss of its good workforces and likewise it will lead to a reduction in output and quality [1].

Essentially, HRM is an all-inclusive set of administrative activities and tasks concerned with developing and maintaining a competent workforce-human resource. HRM intends to enable organizational competitiveness, enrich throughput and superiority, stimulate individual growth and

expansion, and conform through legal and social obligation [2]. In addition, in any establishments, they require to strive meritoriously in terms of budget cost, superiority, service and novelty. Certainly, all the above can be determined by retaining enough right people and offering the right services that can be arranged in the applicable locations at appropriate points in time [2].

The applications of HRM that are within Artificial Intelligent (AI) techniques can be used to help decision makers to solve unstructured decisions. Data Mining is one of AI technologies that have been industrialized to explore and analyze large quantities of data so that to determine meaningful patterns and rules. In actual fact, such data in HRM can supply rich properties for knowledge discovery and decision support tools [3]. There are many areas at the present time such as marketing, finance, telecommunication, manufacturing, medical etc... that have used and adapted data mining techniques [3].

Thus, in order to help companies an early warning system to forecast leaving of their coworkers must be constructed; one must attempt to find out the causes through a feature selection model. Feature selection turned out to be a significant subject in academic fields, such as data mining, machine learning, and pattern recognition [1]. Feature selection which is also called as a subset selection is a process frequently used with machine learning techniques in which subsets of the relevant features obtainable from the data are selected for an application of a learning algorithm. The finest subset comprises the least numbers of dimensions that utmost will give best accuracy. This is an imperative phase of preprocessing and is one of two ways of avoiding the curse of dimensionality and the other is called feature extraction) [4].

Rough Set Theory (RST) is a worthwhile approach to solve problems with uncertainty, imprecision, and incompleteness. This approach can be used to extract fuzzy rules, reasoning with vagueness, classification, feature selection, etc. Combining fuzzy and rough sets is a key direction to tackle uncertainty in real data sets in data mining and machine learning techniques. They do not conflict with each other but rather complement each other in some aspects. These two theories have been used for feature selection and would produce better results, which can be used in classification and prediction at later stages.

In our case study, employee information dataset are used via applying a combined filter feature selection approach for gaining information among the features, then, using Fuzzy Rough Set feature selection approach for capturing the vagueness present in the dataset and selecting the best features which are useful for the classification and prediction tasks. Finally, three classification algorithms are used namely; Fuzzy Rough Nearest Neighbor, Decision Tree, Naïve Bayes.

This paper is organized as follows. In Section 2, we present related works. Section 3, describes the steps of the proposed system, this includes describing each subsection in the proposed system. Then, section 4, describes the classification techniques that are used for learning and evaluating the accuracy of the system. In section 5, results and analysis of using those techniques which are represented with tables. Finally, the key points are outlined and concluded.

II. RELATED WROKS

In the literature several research works were documented on using data mining techniques to solve the HRM problem. The key aspect of data mining is to extract knowledge from a data set and put it in a human-understandable structure [5]. Below, we review some of the techniques and approaches that were used in the above mentioned area.

Hsin-Yun Chang in 2009, suggested a system which represents a new method that could select subsets efficiently. In addition, the author also investigated and discussed the reasons of employers voluntarily turnover so that to increase the classification accuracy and to help managers avert employers' turnover. The mixed feature subset selection was used in this study, combined Taguchi method and Nearest Neighbor classification rules to select feature subset and analyze the factors are implemented to find the best predictor of employer turnover. The results showed that the number of total factors that are important to the employers was 18. In addition, the accuracy of correct selection was 87.85% which was higher than before using this feature subset selection method (80.93%) [1].

Jantan, Hamdan, and Othman in 2009 used decision tree C4.5 classification algorithm to produce the classification rules for Human Talent Prediction in HRM system. Finally, the produced rules are assessed through the unseen data in order to approximate the accuracy of the prediction result. The accuracy for the used classification technique was 95.0847 % [6].

In another research work of the same research team work, in 2011, the authors attempted to use the classifier algorithm C4.5, Random Forest for decision tree, Multilayer Perceptron (MLP) and Radial Basic Function Network (RBF). In the initial stage of this study, the selected classifier algorithms for a sample of employee data were used. In this case, they focused on the accuracy of the techniques to find the suitable classifier for the HRM data. The accuracy for each classification technique was scored as 95.14%, 74.91%, 87.16%, and 91.45% respectively [3].

In 2013, Florence and Savithri in [5], discovered knowledge from talent knowledge acquisition process. In their proposed system, the C4.5 classifier algorithm is used to evaluate the performance of the individuals. This technique was used to construct classification rules to predict the potential talent which would help in determining whether the individual was fit for the assessment or not.

Another work in the literature in 2013 is conducted by Jantawan and Cheng-Fa Tsai [7], they proposed a system to predict whether a graduate has been employed, remains unemployed, or is in an undetermined situation after graduation. This prediction is done by using various algorithms under Bayesian and decision methods to classify a graduate profile as employed, unemployed, or other. Results show that the WAODE algorithm, which is a variant of the Bayesian algorithm, achieved the highest accuracy rate (99.77%). The average accuracy rate of other Tree algorithms was 98.31%.

Throughout our investigation in the mentioned area, we found that there is little or no research works that were previously conducted using fuzzy methods in HRM systems, and since classification of HRM system has not been used in Kurdistan region or at Salahadin University, thus, we found it very important to determine the appropriate system for solving the challenges of Human Resource Talent Management.

In this system supervised classification methods with suitable feature selection techniques are applied for the purpose of providing the best results. Filter feature selection approaches are used (IG, GR, OneR then, they are combined), next, via using Fuzzy Rough Set, the suitable and relevant subset of features is selected to improve the accuracy of classification and obtain better results. In this study three classification techniques are used (Fuzzy Rough NN, Decision Tree, and Naïve Bayes).

III. THE PROPOSED FRAMEWORK

The proposed system for the HRM classification is described in the following subsections:-

A. Data Collection

A common and useful way to collect data can be via a survey approach which has the advantage of being very structured, in addition, it is easily replicable, and possibly to compare the results with surveys that had been previously undertaken. Researchers who are interested in the results are actually not physically close to participants who fill in the survey, thus, it also allows for privacy and anonymity, and facilitate people to respond in a more honest way [8].

Surveys can be carried out by a large number of participants [8]. Thus, because of these advantages of survey approach, the data collection in the research work is done through a survey that was given to participants from companies and organizations in Kurdistan. Table 1 shows relevant variables or attributes for an employee which has two parts, one part is filled by the employee, and the other

part is filled by the director or the supervisor who oversees the employee.

TABLE I. **R**RELEVANT FEATURES AND ATTRIBUTES FOR EMPLOYEE DATA SET.

Filled by employees			
No.	Variable name	No.	Variable name
1	Gender	17	Position
2	Age	18	Department
3	Education background (qualification)	19	Computer skills
4	Language	20	Job security
5	Marriage	21	Smoking
6	Partner working	22	Transportation
7	Number of children	23	Vacation days
8	Average age of their children	24	Nationality
11	Resident	25	Employment type
12	Job time	Filled by Director	
13	Hours of work		
14	Salary	26	Number of activities
15	Years of service	27	Number of penalties
16	Social assurance	28	absence days

B. Data Analysis and Preprocessing:

Data must be cleaned from noise, missing values, and outliers so that to gain better learning and results, thus, data analysis must be conducted for the purpose of detecting errors, as these errors will lead to inconsistency and must be removed. After the questionnaires were collected, the process of preparing the data was accomplished [9]. Then, the types of data have been reviewed and modified. Some attributes like Age and Years of Service are entered in continuous values. So, they are modified and illustrated via ranges. Other attributes like Language is generalized to include fewer discrete values than they already have [9]. So, our dataset contained missing values especially in Salary attribute because respondent did not answer all questions in the questionnaire during the dataset collection, thus, each missing value is replaced with the mean of the attribute, this is how the missing values are treated. Obviously, the mean is calculated according to all known attribute values. This method is convenient with numeric attributes only. So we used it for handling missing values in the Salary attribute.

C. Proposed Algorithm for Feature Selection

The advantages of feature selection process are to improve learning accuracy and to speed up data mining algorithms [10]. Any data in the real world may contain some irrelevant features that have to be discarded. Feature selection is a vital pre-processing stage for choosing a particular subset of attributes from the main data set [11]. The aims of feature selection are as follows: to reduce the dimensionality of the data, to improve the prediction accuracy of the classifier, and to reduce the computational cost. Most of the learning algorithms become computationally inflexible when the number of features is

relatively large. A conventional feature selection process comprises four basic stages (See Figure 1), these are namely; subset generation, subset evaluation, stopping criterion, and result validation [11].

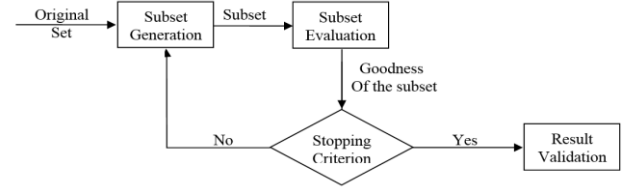


Fig. 1. Four key steps of feature selection.

In the first stage, a search process is conducted to generate subsets which are candidates for evaluation depending on a particular search strategy. Thus, the evaluation for each candidate subset is carried out, and then comparison is conducted against the prior best subset based on a particular evaluation condition. The new subset replaces the prior best subset if in case the new subset is performed better. These two processes are repeated till a particular stopping condition is met. Then after, the validation process on the selected best subset via different data sets is conducted [12].

In this research work, a new proposed algorithm for feature selection is introduced, literally, three filter feature selection methods are used, these are namely; Gain Ratio (GR), Information Gain (IG), and OneR attribute evaluation. The three feature selection methods are combined via applying the Union operation to produce their final results. Then a Fuzzy Rough Subset approach is applied as a feature selection method for evaluating the combined result, in addition the 10 cross validation attribute selection mode is applied. Figure 2 shows the new proposed algorithm for feature selection. Details of the above algorithms are explained in the following subsections.

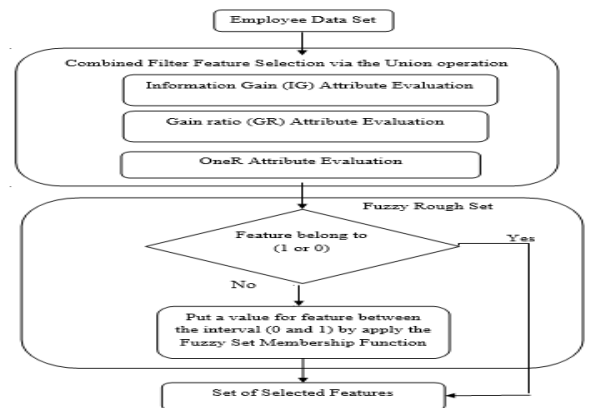


Fig. 2. Proposed algorithm for Feature Selection.

1) Information Gain Attribute Evaluation

(IG) Information gain means the amount of information value that can be provided by each feature [13]. Which is the amount that can be used to determine if the feature is

considered for selection or it can be deleted. A specific threshold value defined for conducting the selection process. If the information gain value for a feature is greater than the specified threshold value, then, the feature is selected [13]. The information gain depends on the entropy concept, it attempts to select attributes that can have large numbers of different values as opposed to attributes with the fewer values, despite the fact that the last is further instructive and useful, this is considered to be the key issue of information gain [14]. The idea of entropy is typically in principles of information gain attribute ranking approaches; it is used to describe the pureness of a random collection of instances. The entropy of Y is expressed as in equation (1) [15]:-

$$H(Y) = -\sum_{y \in Y} P(y) \log_2(P(y)) \quad (1)$$

$P(y)$, represents the minimal probability density function for the variable Y . when the Y observed values in the training data set S are separated with accordance to the values of a second feature X , and the Y with entropy with respect to the separation prompted by X is less than the Y entropy before separation, then, it can be concluded that features Y and X are related. After observing X , the Y entropy can be expressed as in equation (2) [15]:-

$$H(Y/X) = -\sum_{x \in X} P(x) \sum_{y \in Y} P(y/x) \log_2(P(y/x)) \quad (2)$$

$p(y/x)$, represents y conditional probability when x is given. When the entropy is given as a pureness criterion in a training set S , then a measure reflecting additional information can be defined about Y provided by X that represents the amount via which the Y entropy gets reduced. This is called information gain measure as described via equation (3) [15]:-

$$IG = H(Y) - H(Y/X) = H(X) - H(X/Y) \quad (3)$$

Equation (3) is a symmetrical measure of information gain which means that the information gained about Y after X detected is equivalent to the information gained about X after Y is detected. It is noted that the criterion of information gain is inclining for some features that have more values which are not further instructive [15].

2) Gain ratio(GR) attribute evaluation

The aim of the Gain Ratio approach is to integrate split information of features into Information Gain statistic. A gene split information can be achieved through calculating how approximately and homogeneously it splits the data [16]. It is practical that the measure of the information gain is inclined to test with many consequences. Thus, the attributes with a large number of possible values over attributes with a fewer values are chosen despite that the latter is more useful.

If we consider an attribute acts as a unique identifier such as an employee id in the employee dataset. A split on employee id would result in a large number of separations as each record in the database has a unique value for an employee id. So, the information essential to classify database with this separations would be Info employee $ID(D) = 0$. Obviously, this separation is impractical for classification task [14].

The estimated information desirable to classify a given sample can be expressed via equation (4):-

$$I(D) = -\sum_{i=0}^n P_i \log_2(P_i) \quad (4)$$

P_i denotes the probability that an instance be appropriate to class C_i . Assume an attribute A has v distinct values, d_{ij} denotes the number of instances of class C_i in a subset D_j . D_j comprises those instances in D that have value a_j of A . The entropy depended on partitioning into subsets via A can be expressed in equation (5):-

$$E(A) = -\sum_{i=1}^n I(D) \frac{(d_{1i} + d_{2i} + \dots + d_{mi})}{d} \quad (5)$$

Equation (6) describes the encoding information that can be obtained via separating on A .

$$Gain(A) = I(D) - E(D) \quad (6)$$

The gain ratio can be represented by equation (7):-

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad (7)$$

$SplitInfo(A)$ can be expressed via equation (8):-

$$SplitInfo_A(D) = -\sum_{j=1}^v \left(\frac{|D_j|}{|D|} \right) \log_2 \left(\frac{|D_j|}{|D|} \right) \quad (8)$$

Obviously, the splitting attributes is the attribute that has the maximum value of gain ratio [14].

3) OneR Attribute Evaluation

OneR method is another evaluation method for selecting features which is widely used in the data mining application. This approach attempts to construct a rule for each attribute in data set and choose the rule that has the smallest error [17]. In simple way a classification rule is described as $r = (p, c)$ where, p represents a precondition that can perform a sequences of tests that are assessed as true or false and c is a class which can apply samples which covered r , where r represents the rule. Rule based methods can be described by three stages; these are producing rule R on training data S , removing the training data covered by the rule and repeating the process [18].

D. Fuzzy Rough Subset Evaluation

The concept of Fuzzy Rough Set Theory can be constructed based on two other theories that are rough set theory and fuzzy set theory [19]. The two theories are explained as follows:-

1) Rough Set Theory

The idea of Rough Set Theory suggests a fresh mathematical method to imprecise [20]. The core of RST is described via the indiscernibility concept [21]. Assuming $I = (U, A)$ as an information system, U represents a non-empty set of finite sample cases and A is a non-empty finite set of features in a way that $a: U \rightarrow Va$ for every $a \in A$. Va is a set of values that a has [21]. Given the fact that $P \subseteq A$, there is an associated equivalence relation ($IND(P)$) which can be expressed in equation (9) [18]:-

$$IND(P) = \{(x, y) \in U^2 \mid \forall_a \in P, a(x) = a(y)\} \quad (9)$$

Thus, the U partition is produced via $IND(P)$ is expressed via $U/IND(P)$ or U/P and can be premeditated as shown in equation (10) [18]:-

$$U/P = \otimes \{U/IND(\{a\}) : a \in P\} \quad (10)$$

And $A \otimes B$ can be calculated via equation (11):-

$$A \otimes B = \{X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq \emptyset\} \quad (11)$$

When $(x, y) \in IND(P)$, this means that both x and y are unnoticeable via P attributes. The equivalence classes of the P indiscernibility relation can be described as $[x]_P$. Assume that $X \subseteq U$. X is estimated through merely the information confined in P via building both the P-lower and P-upper approximations of X [21]. The P-lower approximation of a set X with respect to P is defined as $\underline{P}(X)$, which includes the set of all objects that can be certainty categorized as members of X with respect to P [20].

$$\underline{P}X = \{x : [x]_P \subseteq X\} \quad (12)$$

The P-upper approximation of a set X with respect to P , is described via by $\overline{P}(X)$, which is the set of all objects which can be only classified as possible members of X with respect to P , see equation (13):-

$$\overline{P}X = \{x : [x]_P \cap X \neq \emptyset\} \quad (13)$$

Assume P and Q be equivalence relations over U , then the three regions such as positive, negative and boundary can be expressed via equations (14), (15), and (16) [18]:-

$$POS_p(Q) = U_{x \in U/Q} \underline{P}X \quad (14)$$

$$NEG_p(Q) = U - U_{x \in U/Q} \overline{P}X \quad (15)$$

$$BND_p(Q) = U_{x \in U/Q} \overline{P}X - U_{x \in U/Q} \underline{P}X \quad (16)$$

The positive region $POS_p(Q)$ comprises all objects of U which is categorized to classes of U/Q via the information in P attributes. $BND_p(Q)$, describes the boundary region which is the set of objects that might possibly, but not definitely be categorized. $NEG_p(Q)$, describes the negative region which is the set of objects that cannot be categorized to classes of the positive region (U/Q) [18].

2) Fuzzy Set Theory

The fuzzy set theory is implemented to detect the imprecision existent in the data sets which would be difficult by using conventional set theory in which elements could belong to either a set or not [18]. This idea is protracted via fuzzy set theory, in which degrees of membership of elements to sets are allowed. Prior to the establishment of the concept of fuzzy set theory, elements would either a membership of 1 or a membership of 0. This limitation is removed via using fuzzy set theory through allowing memberships to take values in between $[0, 1]$. A set of $A = \{x, \mu A(x) \mid x \in U\}$ is a fuzzy set. The function $\mu A(x)$ is the membership function for A , literally, this means representing each element of the universe U to a membership degree in between $[0, 1]$. A normal fuzzy set which includes at least

one element with a membership degree of 1, notice that the universe may be discrete or continuous [18].

3) Fuzzy Rough Set Theory

The main purpose of the fuzzy set theory approach is to create vague information, whereas, the rough set theory approach creates imperfect information. Both approaches are not opposing but are perfecting each other [19]. Therefore, this approach of fuzzy rough set can be described via two fuzzy sets; fuzzy lower and upper approximations, and is obtained by encompassing the conforming crisp rough set notions [21]. The fuzzy P-lower and P-upper approximations are described via equation (17) and (18) [18]:-

$$\mu \underline{P}X(F_i) = \inf_x \max\{1 - \mu F_i(x), \mu X(x)\} \forall i \quad (17)$$

$$\mu \overline{P}X(F_i) = \sup_x \min\{\mu F_i(x), \mu X(x)\} \forall i \quad (18)$$

A fuzzy equivalence class belonging to U/P is described by F_i , keeping in mind that though the universe of discourse in feature selection is predetermined, this can be not the case normally, thus, the employment of \sup and \inf , are observed, these definitions deviate a bit from the crisp upper and lower approximations, this is because the memberships of individual objects to the approximations don't seem to be expressly accessible. The fuzzy lower and upper approximations are described via equations (19) and (20):-

$$\mu \underline{P}X(x) = \inf_{y \in U} \max\{\mu F(y), \mu X(y)\} \quad (19)$$

$$\mu \overline{P}X(x) = \sup_{y \in U} \min\{\mu F(y), \mu X(y)\} \quad (20)$$

Notice that not every value of $y \in U$, has to be taken into account, so, those where their $\mu F(y)$ is nonzero are considered, which y object is a fuzzy member of equivalence class F . Fuzzy rough set can be described via the tuple $\langle PX, \overline{P}X \rangle$.

In the crisp condition, parts that are included to the lower approximation which can have a membership of 1 that are included in the approximated set are with absolute confidence. Within the fuzzy rough case, parts could have a membership within the range $[0, 1]$, therefore, would permit bigger flexibility in handling uncertainty. Fuzzy rough sets summarize the connected yet distinct ideas of unclearness for fuzzy sets and indiscernibility for rough sets, each of which occurs as a results of uncertainty in data [21]. The crisp lower approximation categorized by the subsequent membership as described in equation [18]:-

$$\mu \underline{P}X(x) = \begin{cases} 1, & x \in F, \quad F \subseteq X \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

The case of the fuzzy lower approximation in equation (21) is also rewritten as in equation (22):-

$$\mu \underline{P}X(x) = \inf_{y \in U} \{\mu F(y) \rightarrow \mu X(y)\} \quad (22)$$

IV. CLASSIFICATION AND PREDICTION TECHNIQUES

It is useful to apply a classification model for predicting performance depending on a dataset from certain companies [9]. Selecting useful features in our dataset via using fuzzy Rough Subset Evaluation as a feature selection algorithm for combining the results of three feature selection algorithms is discussed in the previous sections. Then, Fuzzy Rough NN, Decision Tree and Naive Bayes algorithms are used as a classifier in this research work. The details of these algorithms are described in the subsections below.

A. Fuzzy Rough NN

Fuzzy Rough NN algorithm stands for Fuzzy Rough Neural Network (FRNN) which combines fuzzy rough approximations with the classical Fuzzy Neural Network approach. The main concept of this approach is that the lower and upper approximations of a decision class are premeditated via the nearest neighbors of a test y object [22]. The FRNN algorithm first checks the K nearest neighbors (NN) of a desired sample t and then categorizes the desired target instance to the class C in which the sum is maximal, with R a fuzzy indiscernibility function [19], this is expressed in equation (23).

$$(R \downarrow C)(t) + (R \uparrow C)(t) \quad (23)$$

The upper and lower approximations only consider the examples of NN [19] and are expressed as follows:-

$$(R \downarrow C)(t) = \min_{x \in NN} I(R(x, t), C(x)) \quad (24)$$

$$(R \uparrow C)(t) = \max_{x \in NN} T(R(x, t), C(x)) \quad (25)$$

When $(R \downarrow C)(y)$ value is great, then, this means every value of y 's neighbors included in class C . A high value of $(R \uparrow C)$, would state that at least one neighbor included in the class [22].

B. Decision Tree

Decision Trees contains nodes with one node is represented as the root of the tree and the others as the leaves of the tree. The induction of decision tree is thoroughly related to a rule induction, every single path starting from the root of a decision tree to at least one of its leaf is often reworked into a rule through conjoining the tests on the trail to create the antecedent half, and regarding the category prediction of the leaf as the category value [23].

The attribute values of the instance are tested against the decision tree to classify an unknown instance. Thus, *IF-THEN* rules are built for every path right starting from the root all the way down to a leaf node. It is clear that every attribute value pair alongside a specified path can create a conjunction in the rule antecedent or the *IF* part. The leaf node can have the class prediction, would create the rule consequent or the then part [24].

The C4.5 technique is one of the decision tree families that can produce both decision tree and rule sets; and construct a tree for the purpose of improving prediction accuracy. Moreover, C4.5 models are easy to understand as the rules that are derived from the technique have a very straightforward interpretation [6].

C. Naive Bayes classifier:

The Naive Bayes supported rules, which adopts the attributes $X_1 \dots X_n$ all in theory are autonomous of one another, given Y [24-26]. The probabilistic model of naive Thomas Bayes classifier relies on theorem, and also the adjective naive comes from the idea that the attributes of a dataset are reciprocally self-regulating. Assuming a problem of supervised learning within which the approximation of unknown desired function $f : X \rightarrow Y$ is the main objective or evenly $P(Y / X)$. Let Y be a variable of a Boolean random valued, and X could be a vector holding n Boolean features [26-28]. Mathematically speaking, the type of a classifier can be described via the following equation:-

$$f_i(X) = \prod_{j=1}^N P(x_j \cap C_i) P(C_i) \quad (26)$$

Where c_i , $i = 1, 2, \dots, N$, represents potential class labels and $X = (x_1, x_2, \dots, x_N)$ represents a vector of features.

The training stage for learning a classifier can be made up of approximating restricted probabilities $P(x_j c_i)$ and prior probabilities $P(c_i)$. Obviously, $P(c_i)$ can be determined via totaling the training instances that belong to class c_i and then, the resulting count is divided by the training set size.

By the same token, simply spotting the frequency distribution of attribute x_j contained by the training subset that is labeled as class c_i would estimate conditional probabilities.

V. RESULTS AND DISCUSSION

As mentioned, in this research work, three different classification techniques, namely; Fuzzy Rough NN, Decision Tree, Naive Bayes, with ensemble feature selection (three types of feature selection are used, namely; IG, GR, and OneR, in addition, the Fuzzy Rough Set technique is used as an amalgamation) are used. Thus, different experiments with different modeling structures and data sets are implemented to examine the performance of the system. In the feature selection stage, the filter + fuzzy rough set techniques are tested on the data set with 1000 instances of an employee record, each record contained 30 input features and two output labels of classification. Ultimately, number of features is reduced to 11 features, where it was originally 30 features. These 11 selected features are materialized to be very relevant and successfully affected the performance of the overall system during the learning and predicting stages.

Table 2, shows the obtained results from the system on the original data set without feature selection. It can be clearly seen that the best performance is produced by the Decision Tree Technique.

Table 3, shows results on the original data set with feature selection. Obviously the best performance is produced by the Fuzzy Rough NN. It is worth noticing that the behavior of Fuzzy Rough NN is improved due to dimension reduction in the number of features as this reduction in dimension will help reduce the parameters in its structure which leads to furthermore reduction in creating fewer edges and fewer rules.

Table 4, shows the overall accuracy results on different classifiers, it is clear, that the accuracy of both Fuzzy Rough NN and Naïve Bayes are improved using feature selection technique. Whereas the Decision Tree performed slightly less with feature selection.

Table 5, represents the confusion matrix and accuracy measurements for each classifier such as True Positive, False Positive, Precision, Recall, and F-Measure. Where each measurement value extracted from the confusion matrix according to the following equations:-

$$precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (27)$$

$$recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (28)$$

$$F - measure = \frac{2 * precision * recall}{(precision + recall)} \quad (29)$$

VI. CONCLUSION

This paper presented combining Fuzzy Rough Set with Salient Features as a solution for the classification of HRM system. The data set is collected from different corporations in Kurdistan region.

Several experiments are conducted to determine effective feature selection methods, then, the classification task is conducted via Fuzzy Rough Nearest neighbors, Decision Tree and Naïve Bayes. It is noticed that the Fuzzy Rough NN outperformed other algorithms when feature selection is applied. Therefore, using combined feature selection method with Fuzzy Rough Set method is suitable for feature reduction; also using Fuzzy Rough NN method for classification task is convenient and effective to improve the performance of the HRM system.

ACKNOWLEDGMENT

The authors of this paper would like to thank Computer Science and Software Engineering departments, at both Sulaimania and Salahadin universities.

TABLE II. SHOWS DIFFERENT PERFORMANCE RESULTS OF DIFFERENT CLASSIFIERS WITHOUT FEATURE SELECTION.

Classifier	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Relative Absolute error	Root relative squared error	Total Instance
Fuzzy Rough NN	973.3998	26.6002	0.9468	0.0913	0.2082	18.2564%	41.6434%	1000
Decision Tree	978.4557	21.5444	0.9569	0.0287	0.1436	5.7455%	28.7205%	1000
Naive Bayes	888.7	111.3	0.7774	0.132	0.312	26.3899%	62.4054%	1000

TABLE III. SHOWS DIFFERENT PERFORMANCE RESULTS OF DIFFERENT CLASSIFIERS WITH FEATURE SELECTION.

Classifier	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Relative Absolute error	Root relative squared error	Total Instance
Fuzzy Rough NN	981.1736	18.8264	0.961	0.0857	0.1986	17.7445%	40.4155%	1000
Decision Tree	976.0867	23.9133	0.9507	0.0303	0.1468	6.2786%	29.8836%	1000
Naive Bayes	935.9843	64.0157	0.8693	0.0675	0.2	13.971%	40.7105%	1000

TABLE IV. SHOWS ACCURACY RESULTS OF DIFFERENT CLASSIFIERS WITH AND WITHOUT FEATURE SELECTION.

Classifier	With Feature Selection	Without Feature Selection
Fuzzy Rough NN	98.1174%	97.34%
Decision Tree	97.6087%	97.8456%
Naive Bayes	93.5984%	88.87%

TABLE V. CONFUSION MATRIX OF DIFFERENT MEASUREMENT AVERAGE RESULTS FOR EACH CLASSIFIER

CLASSIFIER	TP	FP	Precision	Recall	F-Measure
Fuzzy Rough NN	0.981	0.019	0.981	0.981	0.981
Decision Tree	0.981	0.019	0.981	0.981	0.981
Naive Bayes	0.948	0.052	0.95	0.948	0.948

REFERENCES

- [1] H. Y. CHANG, "Employee Turnover: A Novel Prediction Solution with Effective Feature Selection," *Wseas Transactions On Information Science And Applications*, Issue 3, Volume 6, March 2009.
- [2] H. Jantan, A. R. Hamdan, and Z. Ali Othman, "Knowledge Discovery Techniques for Talent Forecasting in Human Resource Application," *International Scholarly and Scientific Research & Innovation International Science Index*, Vol:3, No:2, 2009.
- [3] H. Jantan, A. R. Hamdan and Z. A. Othman, "Towards applying Data Mining Techniques for Talent Management," *International Conference on Computer Engineering and Applications, IPCSIT vol.2, IACSIT Press, Singapore*, 2011.
- [4] L. Ladha, and T. Deepa, "Feature Selection Methods And Algorithms," *International Journal On Computer Science & Engineering*, Vol. 3 Issue 5, pp. 1787, 2011.
- [5] A. M. Florence.T and R. Savithri, "Talent Knowledge Acquisition Using C4.5 Classification Algorithm," *International Journal Of Emerging Technologies In Computational And Applied Sciences*, pp. 406-410, 2013.
- [6] H. Jantan, A. R. Hamdan and Z. A. Othman, "Human Talent Prediction in HRM using C4.5 Classification Algorithm," *International Journal on Computer Science and Engineering*, Vol. 02, No. 08, 2010.
- [7] B. Jantawan and Cheng-Fa Tsai, "The Application of Data Mining to Build Classification Model for Predicting Graduate Employment," *International Journal of Computer Science and Information Security*, Vol. 11, No. 10, October 2013.
- [8] V. Anderson "Research Methods in Human Resource Management," Published by the CIPD 2009.
- [9] Q. A. Al-Radaideh, E. Al-Nagi, "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance," *International Journal of Advanced Computer Science and Applications*, Vol. 3, No. 2, 2012.
- [10] M. Modi and S. Patel, "An evaluation of filter and wrapper methods for feature selection in classification," *International Journal of Engineering Development and Research*, Volume 2, Issue 2, 2014.
- [11] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *Ieee Transactions on Knowledge And Data Engineering*, Vol. 17, No. 4, April 2005.
- [12] C. S. Yang, and L.Y. Chuang, C. H. Ke, and C. H. Yang, "A Hybrid Feature Selection Method for Microarray Classification," *IAENG International Journal of Computer Science*, 21 August 2008.
- [13] A. Soufi, A. Taleb, A. A. Mohamed, O. A. Mohamed and A. H. Abdelhalim, "Hybridizing Filters and Wrapper Approaches for Improving the Classification Accuracy of Microarray Dataset," *International Journal of Soft Computing and Engineering*, Volume-3, Issue-3, July 2013.
- [14] S. H. Vege, "Ensemble of Feature Selection Techniques for High Dimensional Data," Master thesis, The Faculty of the Department of Mathematics and Computer Science, Western Kentucky University, May 2012.
- [15] J. Novakovic, "Using Information Gain Attribute Evaluation to Classify Sonar Targets," *17th Telecommunications forum TELFOR, Serbia, Belgrade*, November 24-26, 2009.
- [16] A. Hassan, A. S. Abou-Taleb, O. A. Mohamed and A. Hassan, "A Hybrid Feature Selection approach of ensemble multiple Filter methods and wrapper method for Improving the Classification Accuracy of MicroarrayData Set," *International Journal of Computer Science and Information Technology & Security*, Vol. 3, No.2, April 2013.
- [17] J. Novaković, P. Strbac, D. Bulatović, "Toward Optimal Feature Selection Using Ranking Methods And Classification Algorithms," *Yugoslav Journal Of Operations Research*, 21, Number 1, pp. 119-135, March 2011.
- [18] R. Jensen, "Combining rough and fuzzy sets for feature selection", Phd thesis, School of Informatics, University of Edinburgh, 2005.
- [19] N. Verbiest, "Fuzzy Rough and Evolutionary Approaches to Instance Selection", Phd thesis, Faculty of Sciences, Ghent University, March 2014.
- [20] Z. Suraj, "An Introduction to Rough Set Theory and Its Applications," *ICENCO*, Dec, 2004, Cairo, Egypt.
- [21] R. Jensen, N. M. Parthalaian and C. Cornelis, "Feature Grouping-Based Fuzzy-Rough Feature Selection", *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July, 2014, Beijing, China.
- [22] O. Maimon and L. Rokach, "Data Mining and Knowledge Discovery Handbook," Library of Congress Control Number, Springer New York Dordrecht Heidelberg London, Springer Science and Business Media, LLC 2005, 2010.
- [23] U.M. Ashwinkumar and K. R. Anandakumar, "Data Preparation by CFS an Essential approach for decision making using C4.5 for Medical data mining", *International Journal of Software Engineering Research & Practices*, Vol.3, Issue 1, April, 2013.
- [24] G. Kaur and A. Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes," *International Journal of Computer Applications (0975 – 8887)*, Vol. 98, No.22, July 2014.
- [25] T. M. Mitchell, "Generative And Discriminative Classifiers: Naive Bayes and Logistic Regression," *Machine Learning*, Copyright © 2015.
- [26] T. Rashid, "Improvement on Classification Models of Mutiple Classes through Effectual Processes," *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 6, No. 7, 2015.