

## RESEARCH ARTICLE



### OPEN ACCESS

**Received:** 29.11.2020

**Accepted:** 21.01.2021

**Published:** 02.02.2021

**Citation:** Tabaza B, Al-Habashneh O, Lasassmeh O (2021) An adaptive intelligent framework for assessment & selection process in staffing task. Indian Journal of Science and Technology 14(4): 325-334. <https://doi.org/10.17485/IJST/v14i4.2154>

\* **Corresponding author.**

[omarjo@mutah.edu.jo](mailto:omarjo@mutah.edu.jo)

**Funding:** None

**Competing Interests:** None

**Copyright:** © 2021 Tabaza et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](#))

**ISSN**

Print: 0974-6846

Electronic: 0974-5645

# An adaptive intelligent framework for assessment & selection process in staffing task

**Badeea Tabaza<sup>1</sup>, Obadah Al-Habashneh<sup>1</sup>, Omar Lasassmeh<sup>2\*</sup>**

<sup>1</sup> Detp of Information Management System, School of Business, Mutah University

<sup>2</sup> Detp of Computer Information System, Faculty of Information Technology, Mutah University, Jordan

## Abstract

**Objectives:** To enhance the performance of HR's staffing function by providing an intelligent framework that allows convenient assessment and selection procedures. **Methods:** We proposed a new approach that mainly uses Data Mining (DM) and Machine Learning (ML) to develop and train an intelligent framework by learning the behavior of the staffing committee in assessing and selecting applicants for specific job requirements. It utilizes fuzzy logic to mitigate the decision uncertainty and provide an objective mechanism for filtering best-fit applicants' profiles for the next selection phase. The proposed framework was trained on a labeled dataset consisted of (414) CVs. A 5-fold cross-validation method was used to train and evaluate the proposed framework. The highest accuracy achieved was (84%) at k=2; while the lowest accuracy achieved was (71%) at K=1. **Findings:** The accuracy performance is at acceptable levels and can be improved as more data involved in the training process.

**Keywords:** Staffing; data mining; fuzzy logic; machine learning

## 1 Introduction

The contemporary business setting has become highly competitive and volatile. This setting has made the staffing of the right people more critical than ever before. However, most business firms have a limited staffing budget to invest in Human Resources (HR) strategies and functions i.e. staffing. Staffing is an important function of HRs unites that aim to select the most competent and qualified applicants suited for job vacancies. Furthermore, the shrinking job market is linked to the higher volume of job seekers. This context has increased the complexity of the staffing function associated with HRs units. This function requires a precious assessment and selection procedures to produce the shortlisted candidates by matching the job requirements (experience, skills, knowledge, qualifications, etc.) with the applicant's profile<sup>(1)</sup>.

In the staffing process, the task of shortlisting a few numbers of applicant's profiles can be easy; while the task becomes harder when assessing too many applicant's profiles through their written Curriculum Vitals (CVs). Thus, filtering a huge number of CVs becomes a time and cost challenges; further, it needs to be objective and free of prejudice

task as well<sup>(2)</sup>. These challenges can be stepped using an effective intelligent information system that can efficiently support the filtering process, which is unfortunately not available yet<sup>(3)</sup>.

This study attempts to enhance the performance of HR's staffing function by providing an intelligent approach that allows a convenient assessment and selection procedures. This approach aims to handle the contradiction between the panel members and the uncertainty involved in decision making while staffing. The proposed approach mainly uses Data Mining (DM) and Machine Learning (ML) to develop and train an intelligent framework by learning the behavior of the staffing panel members in assessing and selecting applicants for specific job requirements. It utilizes fuzzy logic to mitigate the decision's uncertainty and provide an objective mechanism for ranking the applicants' profiles for the next selection phase.

## 2 Literature Review

According to Han et al. (2011), DM is a field in computer science involving AI and ML and refers to extracting knowledge from huge amounts of raw data, where we can extract an interesting novel pattern from rules that are found in the data. DM algorithms deal only with structured data (database)<sup>(4)</sup>. DM can be used to carry out several functions, such as text categorization, text clustering, information retrieval, information extraction, and summarization<sup>(5,6)</sup>.

Text categorization is the process of document classification into predetermined labels (Ghosh, Roy, & Bandyopadhyay, 2012), while text clustering is the process of dividing a set of documents into similar groups. Unlike text categorization, text clustering analyzes the documents without knowing labels. It is considered a supervised technique while text clustering is an unsupervised technique.

There is confusion between the two fundamental concepts of Information Retrieval (IR) and Information Extraction (IE). IE is the process of detecting and obtaining predefined information from unstructured data, then converting it into structured data automatically<sup>(7,8)</sup>. It is the process of obtaining specific information from unstructured data<sup>(9)</sup>. IR is the process of retrieving a set of documents that match a user's query<sup>(6)</sup>. The Google search engine is an example of an information retrieval system<sup>(9)</sup>.

Jayaraj and Mahalakshmi (2015) proposed two new algorithms. The first one is called Information Retrieval Configuration File (IRCF), which was utilized to extract the required information from resumes. This algorithm had several steps: Identifying the type of document (word, pdf, excel); the configuration file will be created; the document reader will be selected to read the resume line by line. The problem with this algorithm is that for every condition, the document reader will read the configuration file to extract the required information<sup>(10)</sup>. Let us consider that there are seven conditions, so that process will be repeated seven times for every resume, and this will take a long time.

The second algorithm, called weighted ranking, was employed to rank the resumes. The algorithm ranked the resumes based on the candidate's education level, technical Skills, general Skills, experience and age. Actual Resume Relevancy (ARR) was used to evaluate the performance of an IR system, where ARR gave results of 69.6%. This study did not provide any performance measurement for information extraction processes such as precision or recall. According to Jayaraj,(2015), this assumes that the total years of experience are found in the resumes. However, in some resumes, the candidate does not write this, so we have to take into account the calculation of the candidate's total years of experience.

Yu, Guan and Zhou (2005) proposed a new hybrid model to extract information from Chinese resumes. This model had two steps: the first step was dividing resumes into several sections (blocks) using the Hidden Markov Model (HMM). The second step was the IE from resumes, where HMM was used to extract educational information such as graduate school, degree and major, while Support Vector Machines (SVM) were used to extract personal information such as name, birthday, address, phone, mobile, and email. Also, it was used to enhance the process of personal information extraction. The experimental results carried out on 1200 Chinese resumes showed that the precision was 86% and recall 76% of the personal information and showed the precision of 70% and recall 76% of the educational information<sup>(11)</sup>.

Kopparapu (2010) proposed a new system to extract information from resumes automatically. Also, the system provided a search engine for resumes. Where regular expression technique was used to extract six fields that include qualification, experience, skills, age, name and email. The experimental results carried out in 100 resumes showed that the precision was 87% and recall 71 %<sup>(12)</sup>.

Jiang, Zhang, Xiao and Lin (2009) proposed a new model to extract 18 different pieces of information from Chinese resumes, where resumes were divided into two main blocks: the basic information and complex information. The basic information includes name, gender, birthday, address, mobile phone, email, while complex information includes education, work experience, project experience, award and skills. The regular expression technique was used to extract the basic information where the accuracy was 87%, while SVM and regular expression techniques were used to extract the complex information where the accuracy was 81%. The accuracy of the whole model was 84%<sup>(13)</sup>.

Chuang, Ming, Guang, Bo, and Zhiqing, (2009) proposed a new system to extract information from Chinese resumes. The system had three modules: segmentation module, information extraction module, and feedback control module. The system

divided resumes into small classes using a segmentation algorithm to prevent information overlapping. Then, SVM and regular expression were used to extract the required information from each class. The experimental results carried out on 5000 resumes, where 2000 resumes were used as training data and 3000 were used as a test sample, showed that the accuracy was 84.65%. Should mention here that the system repeats the previous steps to extract the required information from each class and that takes considerable time<sup>(11)</sup>.

### 3 Problem Statement

The staffing process has become more important in the last decades as the business environment became volatile and highly competitive, which has made HR to become more valuable for organizations. The processes of assessing and selecting candidates for a job position are the key elements in staffing decisions. They allow producing the shortlisted candidates by matching their profiles with the job requirements such as experience, skills, knowledge, qualifications, etc. However, performing this task with a huge number of CVs is a time, cost and objectivity issue that needs to be solved by initiative solutions based on intelligent information systems. Unfortunately, HRs still lack such a system to manage the process of staffing to save their time and cost and avoid the panel members' contradiction, uncertainty, and subjective decisions. Therefore, this paper fills the technological gap in HRs staffing function and provides an intelligent approach to enhance the performance of HR's staffing function.

### 4 Aims and Objectives

This study aims to improve the performance of HRs staffing function by providing convenient assessment and selection procedures. It adopts a fuzzy-based adaptive intelligent framework for shortlisting candidates based on job specifications. This requires satisfying the following objectives :

1. Exploring the current methods, techniques, tools and issues in the assessment and selection process.
2. Providing a labeled dataset has been created for this purpose and could be used by other researchers who are interested in the same field in the future.
3. Proposing a fuzzy-based intelligent adaptive approach to improve the assessment and selection of the right candidates and the way of filtering the CVs in the organization.
4. Evaluating the accuracy of the proposed approach.

### 5 Research Questions

This research is carried out to answer the following questions:

1. What are the current solutions used for supporting the process of assessment and selection in staffing?
2. How can an intelligent fuzzy-based framework enhance the assessment and selection process?
3. What is the performance accuracy of the proposed framework?

### 6 Scope and limitations

The scope of the current research paper is limited to the following:

1. It uses open-source and freely available technologies.
2. It focuses on improving the assessment and selection processes for staffing by smart filtering the CVs in an organization.
3. It handles CVs written in the English language only.

### 7 Methodology

In order to achieve the aim and objectives of the research, a multi-step methodology has been applied. It exploits the ML approach where the proposed model is iteratively developed and trained. Thus, the methodology involved five steps to build the proposed approach, which are: problem identification and definition; proposing the approach to address the research problem; collecting data; and training and validating the proposed approach; evaluating the accuracy of the proposed approach. The proposed framework is an adaptive intelligent recommended approach consists of two main levels as shown in [Figure 1](#) .

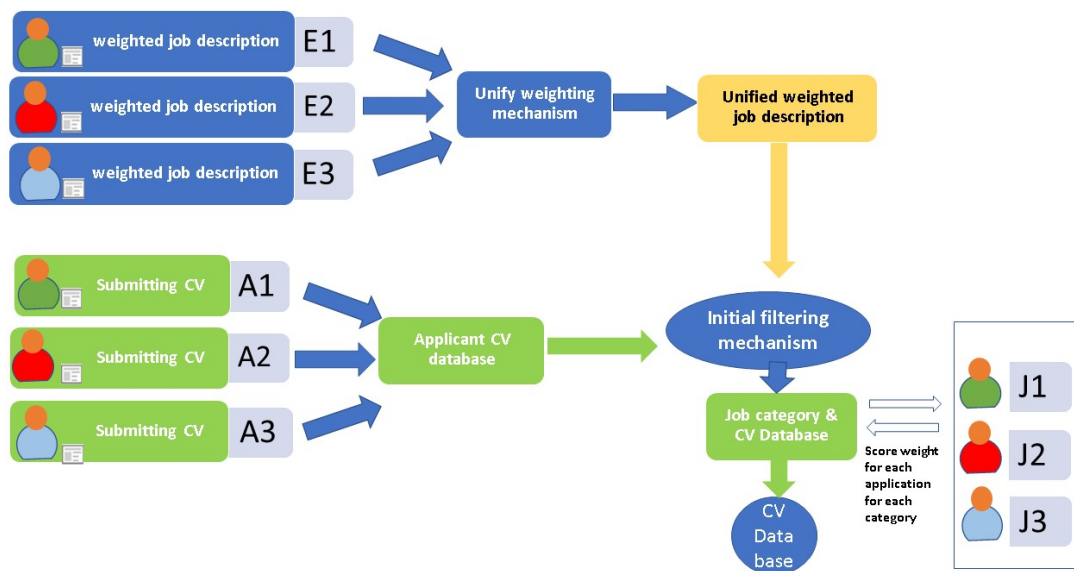


Fig 1. The proposed approach

## 7.1 The first level

### Unified weighting mechanism

Job description usually contains a different job category which had been the different weight of importance between panel members and the job itself. The assessment panel members define the job categories and the weights out of the full mark on the job as per appendix. The unified weighting mechanism proposed in this framework provides a means to resolve this potential contradiction by taking the average weight of the giving weights for each job spec.

### The fuzzy-based contradiction resolution mechanism

As the CVs are scored by different people there a potential contradiction between the member's scores for each CV. In order to handle this contradiction a contradiction fuzzy-based mechanism was developed as follows:

1. Each assessment panel is asked to provide a score for each applicant for each job category.
2. For each panel member, the scores are linguistically labeled using fuzzy sets as shown in Figure 2.

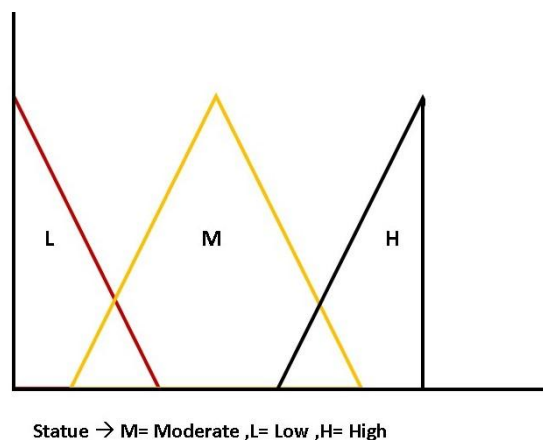


Fig 2. Linguistic labeled using fuzzy set

The linguistic labels of all staffing committee members for each candidate are evaluated against a common-sense fuzzy rule set to identify the final label (L, M, H) for each candidate for each job category. The rule set is shown in Table 1 below.

**Table 1.** The rule set as per common sense

E1*	E2**	E3***	→	Final decision
L	L	L	→	L
L	L	M	→	L
L	L	H	→	M/L
L	M	L	→	L
L	H	L	→	M/L
L	M	M	→	M
L	H	H	→	H
L	M	H	→	M
L	H	M	→	M
M	M	M	→	M
M	M	L	→	M
M	M	H	→	M
M	L	M	→	M
M	H	M	→	M/H
M	L	L	→	L
M	H	H	→	H
M	H	L	→	M/H
M	L	H	→	M/H
H	H	H	→	H
H	H	L	→	H
H	H	M	→	H
H	L	H	→	H
H	M	H	→	H
H	M	L	→	M/H
H	L	M	→	M/H
H	M	M	→	M
H	L	L	→	M/L

\*E1 =Assessment panel member 1

\*\*E2 =Assessment panel member 2

\*\*\*E3=Assessment panel member 3

These results of linguistic label representation are the candidate's scores and the final judgment of panel members. The proposed framework was constructed using filtering the dataset of CV to have the nominated CV for an interview. In the dataset, each candidate is represented as a set of weighted terms which represent their strength compared with other candidates. More precisely, the following steps were followed to develop the dataset:

1. A set of job description JR which led to determine the job requirements were selected.
2. A set of applicants A were selected and for each applicant  $A_n \in A$ .
3. After identifying the sets  $\Omega A_n$  in step 3, the CV information in the dataset was pre-processed and transformed into a set of candidates through a weighting system to define the strength of his CV.
4. The frequency measures: first status (include the resident place), second status (including the year of experience), third status (including knowledge and competencies), fourth status (including skills and abilities) of each candidate term were calculated based on each set  $\Omega A_n$  and used as inputs to a fuzzy system. These frequency measures were used to calculate the term frequency in CVs collection.
5. The clear values of the four input variables (first status, second status, third status and fourth status) were mapped to sets of predefined fuzzy sets which have three linguistic labels (High, Moderate, Low) as per Figure 3. On the other hand; the

output variable Priority has four fuzzy sets associated with four linguistic labels (First, Second, Third, Not Eligible) as shown in Figure 4.

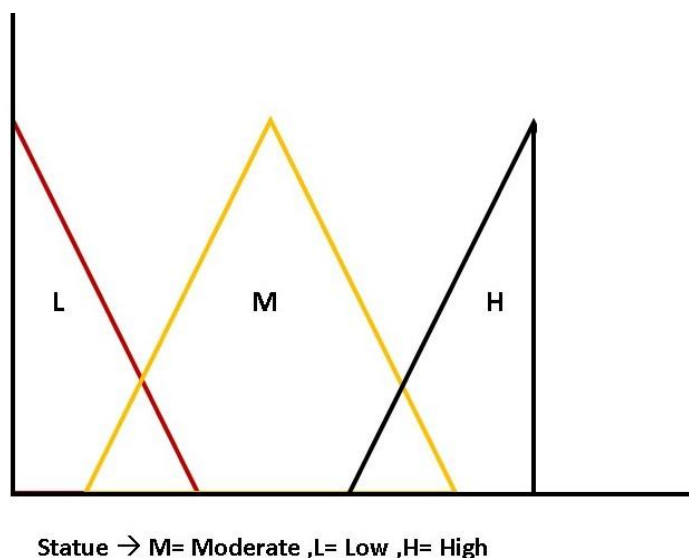


Fig 3. Input: Status of predefined fuzzy sets

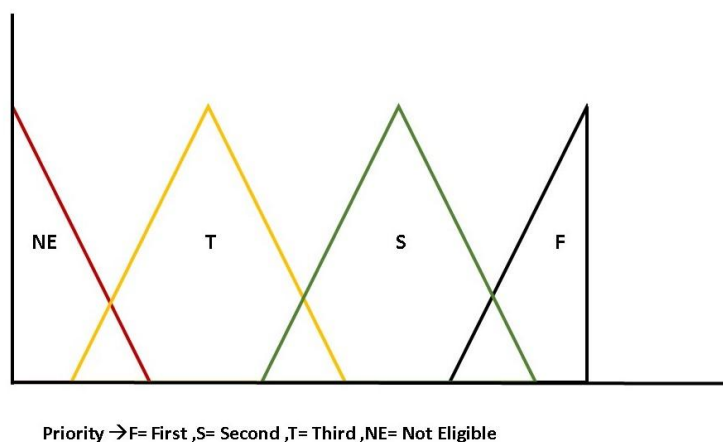


Fig 4. Output variable priority

Member function center of gravity has been used to fuzzified the value to calculate the linguistic label.

#### The Initial List of Candidates

Some job category builds on numeric as “years of experience” which is easy to match the score for it. However, in some other job categories, the value accepted for it is “text” which means the need to use retrieving and matching the keywords.

In order to retrieve information from a candidate's CV the framework proposed the following six steps to retrieve the information needed and save it in a separate dataset and this done by a system:

1. Scan the database to check the presence of the skills and abilities of information fields.
2. Mark the answer as per the number of words specified appearance on the applicant's answer.
3. Store a set of keywords as defined by the recruiter on the database.
4. Retrieve the stored keywords from the database and sets them in a list.
5. Find common words between two lists or arrays using the function intersect.
6. Save results in a database.



### 7.1.1 The second level

#### Fuzzy rules extraction

Let's say that  $A$  is the set of applicants in the data set that contains  $H$  category from CVs, then  $A_h$  is the applicants who apply for a job where  $h=1$  to  $H$ . Each  $A_h$  is associated with the job requirements  $JR$ .  $JR$  consists of  $S$  category where  $JR_s$  is the query category in  $JR$ , and  $s=1$  to  $S$ . Then each  $JR_s$  is associated with its weight  $S$  with each of the lines on dataset  $A_h$  and  $JR_s$  that were computed. These four weights are associated with the predicted relevance  $R_h$ . As a result, each term  $S_z$  is represented as a set of one value  $\{S_1, S_2, S_3, S_4, R_h\}$ . If we consider the four first weights as inputs and the  $R_h$  as a result, then we have a sequence of four input values and one result  $\{S_1, S_2, S_3, S_4, \rightarrow R_h\}$  for each instance in the dataset.

The inputs and output values were mapped to predefined fuzzy sets with the linguistic labels “Low”, “Medium” and “High” based on the Mendel Wang method described in (Wu, Mendel, and Joo, 2010). As shown in Figure 1, the proposed approach, the shapes of the membership functions for each fuzzy set were based on triangle MFs. The triangular MF is specified by three parameters  $\{a, b, c\}$  as in equation 1, which illustrates a triangular MFs defined by triangle for the fuzzy sets<sup>(14)</sup>.

$$\text{Triangle}(x, a, b, c) = \left\{ 0, x \leq a \frac{x-a}{b-a}, a \leq x \leq b \frac{c-x}{c-b}, b \leq x \leq c 0, c \leq x \right\} \quad (1)$$

Where the parameters  $\{a, b, c\}$  (with  $a < b < c$ ) determine the  $x$  coordinates of the three corners of the underlying triangular MF.

The outcome of this step was a set of antecedents and consequences also called ‘if-then’ fuzzy rules where each of the inputs represented by the associated linguistic label {“Low”, “Moderate”, “High”} and outputs represented by the associated linguistic label as shows in Figure 3, If  $B$  is the linguistic label {“First”, “Second”, “Third”, “Not Eligible”} then the fuzzy rule  $BR_h$  is:  $B(S_1), B(S_2), B(S_3), B(S_4) \rightarrow B(R_h)$

#### Framework training and best fuzzy rules set selection

A K-fold of five folds ( $k=5$ ) training and validation method was applied in order to select the rule set that achieves a high level of accuracy in classifying the relevance between the categories of job requirements and the applicant's answers. This method included the following substeps:

1. Dataset partitioning: The fuzzy rules set which resulted from 3.3.2.1 “Fuzzy rule extraction” was partitioned into five equal-sized folds and for each fold will repeat steps 2, 3, 4 were carried out.
2. Training rules set selection: The training rules set was selected by holding out the current fold ( $k_j$ ) The validation process was carried out in  $k$  iterations and in each iteration  $j$ : 1 to  $k$ , the subset  $k_j$  was held out and called hold-out set  $D_h$ . The reset of the subsets was grouped in a training set  $D_t = D - k_j$ .
3. Compression of Fuzzy Rules: A rule compression was performed on the fuzzy rules in the training set  $D_t$ . This was done in order to extract those rules with the maximum firing strength. The rule compression technique was adopted from Renkas, & Niewiadomski (2010), and used later by Doctor and Iqbal, 2014<sup>(15)</sup>,<sup>(16)</sup>. It was based on two quality measures, namely **generality** and **reliability**, for each unique rule pattern. **Generality** refers to the number of instances representing the rule pattern. Reliability reflects the confidence level in the rule pattern. Both generality and reliability were used to calculate the scaled weight of each unique rule pattern. The rule of generality was measured using scaled fuzzy support. The scaled fuzzy support was a frequency ratio of a unique rule pattern in a set of rules having the same consequent as shown in equation 2 and was based on the calculation described in<sup>(17)</sup>. In the proposed approach, it was used to identify and eliminate duplicate instances by compressing the rule base into a set of  $M$  unique and rules modeling the data.

$$\text{Sup}(FR_i) = \frac{CoFR_i}{CoFR_i + CoFR_a} \quad (17, \text{eq.2})$$

Where  $i=1$  to  $M$ ,  $i$  is the index of the rule  $FR_i$  is a unique antecedent combination associated with the consequent linguistic label  $B$  (unique rule pattern) and  $CoFR_i$  is the number of instances that support the rule pattern  $FR_i$  in the data set.  $FR_a$  is the set of the other antecedent's combination, which are different from  $FR_i$  but have the same consequent as of  $FR_i$ .  $CoFR_a$  is the number of instances that support these other combinations  $FR_a$ .

The confidence of a rule is a measure of a rule's validity representing the strength of a unique rule pattern against contradictory rule patterns  $FR_b$ , which refers to the other rule patterns which have the same antecedent combination but with different consequent. Scaled rule confidence is calculated by taking the ratio between the number of instances representing a unique rule pattern  $CoFR_i$  and the number of instances representing the contradictory rule patterns  $CoFR_b$ . Equation 3 shows

the equation of confidence (Steiger & Steiger, 2004)<sup>(18)</sup>.

$$Conf(FRi) = \frac{C_{OF Ri}}{C_{OF Ri} + C_{OF Rb}} \quad (18, \text{eq.3})$$

Calculation of Scaled Rule Weights: the product of the scaled fuzzy support and confidence of a rule was used to calculate the rule's scaled fuzzy weight as shown in equation 4. In this step, the unique rule patterns resulting from the previous step were weighted by calculating a scaled fuzzy weight for each of the patterns. The scaled fuzzy weight was calculated as a multiplication of the scaled fuzzy support and the scaled confidence.

$$SW = Sup \times Conf \quad (18, \text{eq.4})$$

The scaled fuzzy weight **SW** was used to measure how well the rule can represent the data. It was used to rank the fuzzy rule patterns in order to select the most representative rules as described in<sup>(17)</sup>. The scaled fuzzy rule weight was used to select rule patterns with the highest weights out of the other contradictory patterns. So as per checking the dataset and how much is reflected in these rules, we found that 304 CV from 362 CV matching the rules which mean 83.98%.

## 8 Results

The developed fuzzy rule-based system was validated using a well-known validation method called k-fold cross-validation<sup>(19)</sup>. In k-fold cross-validation, the dataset D is divided into k equal size of size h items) subsets called folds. The validation process is carried out for k iterations and in each iteration j: 1 to k, the subset kj is held out and called hold-out set Dh. The rest of the subsets are grouped in a training set Dt = D – kj. The accuracy of the model for each fold ki was calculated using the equation shown in the equation.

$$ACC = \frac{1}{h} \sum_{j=1}^k acc_j \quad (19, \text{eq.5})$$

In order to train and validate the fuzzy rule-based system, a 5-fold cross-validation was applied. The dataset was partitioned into five-folds, representing 20% of the dataset. And then at each iteration, one of these subsets was held out and the system was trained on the other folds representing the remaining 80% of the dataset to extract a set of weighted fuzzy rules. The resulting fuzzy rules of the training process were used to build a fuzzy system that classifies the relevancy of each instance in the hold-out data. The resulting relevancy classifications were compared with the associated linguistic labels of the predicted documented relevancy values from the linear predictive model. Figure 5 presents a sample of the resulting rule of the training process of the rule summarization component for the first fold (K=1). Figure 5 shows a part of the resulted rules from the training process for (k=1)

	S1	S2	S3	S4	Rh	Firing Strength
1	Moderate	High	low	moderate	Second	0.1875
2	Moderate	High	moderate	low	Third	0.2116
3	Moderate	Low	high	low	Third	0.2449
4	Moderate	Low	high	moderate	Second	0.2422
5	Moderate	High	moderate	high	First	0.2400
6	Low	Low	high	low	not	0.2449
7	Moderate	Low	low	moderate	Third	0.2475
8	Moderate	Low	low	high	Third	0.2449
9	High	Low	high	high	First	0.2500

Fig 5. Apart from the resulted rules from the training process (k=1).



In this method a dataset  $\mathbf{D}$  is divided into two subsets;  $\mathbf{D}_t$  (usually 80% of  $\mathbf{D}$ ) and  $\mathbf{D}_h$  (usually 20% of  $\mathbf{D}$ ). Set  $\mathbf{D}_t$ , named as the training set, was used to train the model, while  $\mathbf{D}_h$ , called the testing set or hold-out set, is used to test the model in order to calculate the accuracy.

In order to train and validate the fuzzy rule-based system, a 5-fold cross-validation was applied. The dataset was partitioned into five-folds, representing 20% of the dataset. And then at each iteration, one of these subsets was held out and the system was trained on the other folds representing the remaining 80% of the dataset to extract a set of weighted fuzzy rules. The resulting fuzzy rules of the training process were used to build a fuzzy system that classifies the relevancy of each instance in the hold-out data. The resulting relevancy classifications were compared with the associated linguistic labels of the predicted the priorities of applicants from the linear predictive model as shown in Figure 6.

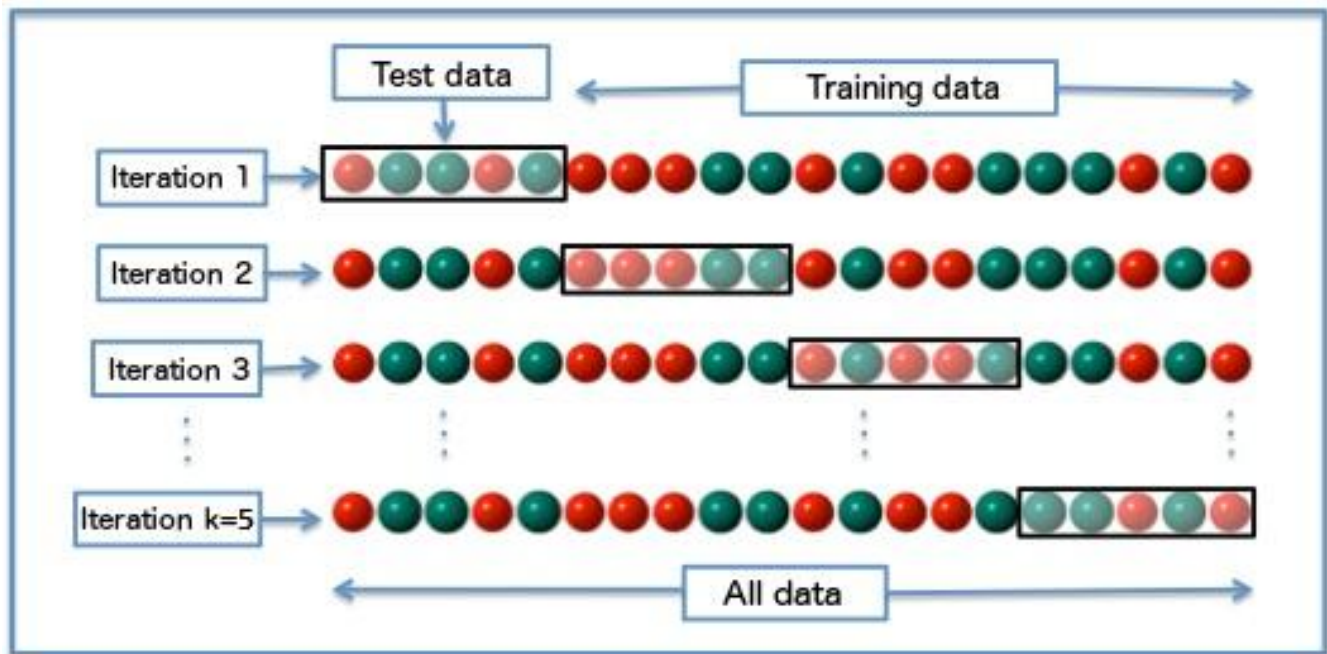


Fig 6. How K folds works

The resulting fuzzy-based approach includes (76) rules to be applied during the testing phase. The dataset contained 414 CVs, which mean 83 CVs on each k; so, the process was repeated five times for the five different folds, and for each fold, the accuracy was calculated for the classifier. Table 2 presents the accuracy for each fold. It can be seen from the table that the highest (84%) accuracy was in the fold K=2 and the lowest accuracy (71%) was in the fold five at K=1. However, the average accuracy of the approach is 77.6 % which was relatively well as the initial performance, hence it was expected to improve as data was stored from the different applicants and different jobs.

Table 2. The accuracy for each fold

Fold K	Accuracy
1	71%
2	84%
3	79%
4	73%
5	81%
Average	77.6%

Results in Table 2 were produced based on the defined fuzzy rules, which applied to the proposed framework to classify the potential candidates based on the following priorities: First, Second, Third, and Not eligible.

## 9 Conclusion

The proposed framework is an adaptive fuzzy based intelligent system, which proves its ability to filter out the best fit candidate using ML approaches. It improves the HR staffing task in terms of assessing and selecting the best-fit candidates using a dynamic weighting schema to enhance the classification accuracy. It consists of five components: unified the weighting approach for each job category, extract information from the dataset, applying text mining technique, fuzzy rules extraction, and best fuzzy rules set selection. In comparison with other approaches presents by other researchers, this approach presents a fuzzy-based approach while other approaches such Jayaraj and Mahalakshmi, Yu, Guan and Zhou, and Saxena adapted other classifiers like NLP, SVM, IRCF, ARR and LP. The proposed framework accuracy reached 77.6%; while a significant number of researches presented frameworks with accuracy between 71% and 87%. They generally adapted how to extract information from CVs with several techniques for doing so; however, this research uniquely extended to evaluate the information and how it is useful for HRs especially in the assessment and selection for staffing.

For future work, this research opens up the opportunities to conduct new researches including more job descriptions, as this study was designed based on the framework of administration, accountant, and IT job description.

## References

- 1) Lu X, Colbry D, Jain AK. Three-dimensional model based face recognition. In: and others, editor. Proceedings of the 17th International Conference on Pattern Recognition; vol. 1. 2004;p. 362–366.
- 2) Alhabashneh OYA, Coventry University. An Adaptive Fuzzy Based Recommender System For Enterprise Search. 2015.
- 3) Saxena C. Enhancing Productivity of Recruitment Process Using Data mining & Text Mining Tools. 2011.
- 4) Han J, Kamber M. Data mining concepts and techniques San Francisco Moraga Kaufman. 2001.
- 5) Feldman R, Sanger J. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge university press. 2007.
- 6) Bazylak J, Weiss PE. Online Evolution: Advantages and Challenges of Online Course Components. *Proceedings of the Canadian Engineering Education Association (CEEAA)*. 2018. Available from: <https://dx.doi.org/10.24908/pceea.v0i0.10182>.
- 7) Chambers N, Jurafsky D. Template-based information extraction without the templates. In: and others, editor. Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies. 2011;p. 976–986.
- 8) Saggion H, Poibeau T. Automatic text summarization: Past, present and future, in Multi-source, multilingual information extraction and summarization, Springer. Springer. 2013.
- 9) Ghosh S, Roy S, Bandyopadhyay SK. A tutorial review on Text Mining Algorithms. *Int J Adv Res Comput Commun Eng*. 2012;1(4):7.
- 10) Jayaraj V, Mahalakshmi V. Augmenting Efficiency of Recruitment Process using IRCF text mining Algorithm. *Indian Journal of Science and Technology*. 2015;8(16). Available from: <https://dx.doi.org/10.17485/ijst/2015/v8i16/53381>.
- 11) Chuang Z, Ming W, Guang LC, Bo X, Zhi-Qing L. Resume parser: Semi-structured chinese document analysis. In: and others, editor. 2009 WRI World Congress on Computer Science and Information Engineering; vol. 5. 2009;p. 12–16.
- 12) Koppurapu SK. Automatic extraction of usable information from unstructured resumes to aid search. *2010 IEEE International Conference on Progress in Informatics and Computing*. 2010;1:99–103.
- 13) Jiang Z, Zhang C, Xiao B, Lin Z. Research and implementation of intelligent chinese resume parsing. *2009 WRI International Conference on Communications and Mobile Computing*. 2009;3:588–593.
- 14) Anand MCJ, Bharatraj J. Theory of triangular fuzzy number. *Proc NCATM*. 2017;p. 80.
- 15) Renkas K, Niewiadomski A. Learning rules for hierarchical fuzzy logic systems using Wu & mendel IF-THEN rules quality measures. *International Conference on Artificial Intelligence and Soft Computing*. 2016;p. 299–310.
- 16) Doctor F, Iqbal R. An intelligent framework for monitoring student performance using fuzzy rule-based linguistic summarisation. *2012 IEEE International Conference on Fuzzy Systems*. 2012;p. 1–8.
- 17) Ishibuchi H, Yamamoto T. Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*. 2005;13(4):428–435. Available from: <https://dx.doi.org/10.1109/tfuzz.2004.841738>.
- 18) Steiger NM, Steiger DM. Knowledge Mining in DSS Model Analysis. *Organ Data Min Leveraging Enterp Data Resour Optim Perform*. 2004;p. 157–169.
- 19) Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Statistics Surveys*. 2010;4(0):40–79. Available from: <https://dx.doi.org/10.1214/09-ss054>.