



# Machine Learning-based Model for Talented Students Identification

Abdelaziz A. Abdelhamid<sup>1,2</sup>

<sup>1</sup>Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra 11961, Saudi Arabia

<sup>2</sup>Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt  
Emails: [abdelaziz@su.edu.sa](mailto:abdelaziz@su.edu.sa)

## Abstract

Identifying the talented university students plays an important role in higher education. Special curriculum can be developed for these students as an outcome from the identification process. This curriculum can be compacted, clustered, and accelerated to match and exploit students' abilities. Current methods for identifying talented students are based on simple identification test in the form of a questionnaire, which is developed for specific age. However, this method of identification cannot cover all aspects of student abilities and inaccurate as it not an iterative process. In this paper, a machine learning approach is proposed for identifying talented students based on their academic performance, which is evaluated repeatedly through their study. In this approach, we measure a set of features representing student abilities, then cluster them based on their features similarity. The proposed approach is applied on a set of 100 university students and shows promising results in identifying the talented group. To emphasize their talent, this group is guided to participate in national competitions that match their abilities, and they could achieve significant ranks.

**Keywords:** Talented student; Academic abilities; Intelligent Agent; Machine Learning.

## 1. Introduction

The formal education in most universities depends mainly on the assumption that academic advisors guide their students to register in courses that have a list completed prerequisites without taking into consideration students' abilities, which can result in learning difficulties for those students through their study. In addition, this formal method does not take into consideration the real existence of differences in students' intellectual, emotional, and social skills. Moreover, talented students are recognized as an important part of the education process [1].

As the occurrence of talented students in the university community is low, the talented education is only provided to a limited set of students. Nevertheless, there are several challenges forming a barrier to the talented students, such as the homogenous education and the negative attitudes from others who are not talented. However, due to the increased knowledge and awareness of these talented students, these negative attitudes are slowly changing. In addition, parents of talented students became keen to get a special education curriculum for their children. In addition, this requires a special care from educational institutions [2].

The identification of talented students is the first and most significant step required for providing talented education. The traditional identification process is based on a paper-based test to measure student abilities through solving complex and difficult problems. However, the immature intellectual ability of students might now allow them to pass this test [3].

To avoid this, a new approach is developed for extracting features that measure the implicit capabilities of talented students based on their academic records, then built a machine learning model for distinguishing talented students from other students using K-means clustering technique. This model could categories students based on their fields of talent. Consequently, we suggest an educational program that matches and considers the type of talent.

During the past decades there were a continuous debate about the identification of superior and talented individuals. Although authors in [4] deduced that the identification of potential talented students cannot be measured using a single number corresponding a domain of talent, other authors in [4, 6] suggested policies and strategies for identifying these students based on ongoing assessment and contextual factors. On the other hand, authors in [7, 8] suggested that the identification of talented students could be done based on a multidimensional approach including the academic performance along with extensively researched and standardized tests.

Different research instruments and techniques have been used by the numerous talent search programs developed in many countries. Authors in [9, 10] indicated that the use of standardized assessments that have been thoroughly studied would often include successful procedures to recognize talented students. Nevertheless, for a good identification process, authors in [10, 11] suggested taking into consideration a set of recommendations for achieving a successful identification, such as considering multiple manifestations of talent, repeat assessment over time to identify additional talented students.

Several studies have shown that among students who do not perform well in exams, talented students can be found [8, 9]. These students are classified into a category known as underachievers. If not classified, they would be a loss to society and the nation that can benefit from their current capabilities. Under-reach is defined as a gap between the academic performance of the student and some index of his or her real capacity, such as intelligence [10]. As such, the academic institution will not be able to recognize talented students by excluding non-performing students and will thus not be able to provide the required learning conditions that will maximize their full potential.

There have been many attempts within the national school system of Malaysia [11] to provide educational programs that appeal to highly competitive students. Students who were eligible to participate in this program had to demonstrate outstanding performance in education. Not so much of an acceleration program was this program, but it was still grade skipping. Students missed a learning year and joined their seniors who were a year ahead of them. For some unexplained cause, this software was in operation for a while before it was put to a halt. Students had to sit for a competency test called Level One Evaluation (Penilaian Tahap Satu-PTS) in year two. Those who passed PTS were offered to bypass year three and move to year four at a certain degree of competency. PTS was abolished in 2002 [12], much like the express class scheme. One of the setbacks was that parents wanted to coach their children to answer PTS-like questions. Although conducting those two training services to support students of high ability, the ministry did not note that they were simply an alternative to the talented education program.

In this paper, we propose a new approach for identifying the talented students based on their academic record. The proposed approach is based on extracting features representing various skills and talent from the academic records. Then, a machine learning technique is applied to cluster and classify the talented students. These clusters of students can be offered a special curriculum that can speed up their graduation.

This paper is organized as follows. Section 2 provides the literature survey and background of identification systems of talented students. In section 3, the proposed identification approach is presented along with the discussion of its benefits in predicting the implicit capabilities of students. Finally, the conclusion and future perspectives are discussed in section 4.

## **2. Related Work**

Talented learners are born with exceptional skills. The skills apply to the naturally inherited characteristics associated with genetic kinship [12]. The inherent abilities are in multiple domains and can be seen from actions and behaviours of high performance, particularly those involving cognitive skills. These include reasoning, planning and problem solving [13] skills. The speed of learning and productivity in the management of the knowledge collected indicate strong cognitive abilities.

A person is talented when he has specific cognitive skills and is able to use them skilfully, as described by [14]. In short, those who can be productive and competitive are talented learners. It is consistent with the revised Gagn's Differentiated Talent and Ability Model, which describes that the natural intellect of individuals is created entirely by catalytic factors (environmental and intrapersonal) that will make the individual succeed and be recognized as talented [15].

However, owing to catalytic causes that do not support the developmental needs of their learning, there are many cases where talented learners are obstructed or unable to demonstrate their innate ability. It suggests that talented qualities can only be seen when they have a high degree of intrapersonal knowledge and environmental factors that facilitate the growth of their potential intellect. Authors in [16] proposed that, as if enrichment, acceleration, and curriculum compression, the process of talented growth takes place in demanding learning approaches. The DMGT was then reinforced by [17] and was designated as the Extended Concept of Talent Creation (EMTD) for Gagne. By entering two developmental models of Gagné, the DMGT and Gagné's Developmental Model for Natural Ability (DMNA), which involves biological basement (genetics) as the starting factor of talented, EMTD is under and above the DMGT theory of talent growth.

The other model by [18] that stated that the word talented of a person is focused on three key characteristics that are highly above average skill, imagination, and job dedication is the Three-Ring Conception of Talent model. The model illustrates that the role of talent is not set or immutable and is continually evolving according to the context that leads to the growth of the person. In his studies, on the other hand, Gardner's Multiple Intelligence Theory [19] showed that the creation of individuals is distinguished through nine main fields, which are verbal-linguistic, mathematical-logical, harmonic, visual-spatial, kinaesthetic, intrapersonal, interpersonal, naturalistic, and existential.

Malaysia has developed a PERMATA program focused on the concept of models and talented theory described above and has developed a talented concept that is in line with the aim of National Education Philosophy to shape the personalities of outstanding students. One of the eight PERMATA programs with the goal of creating potential Islamic scholars is the PERMATA Insan program. The second institution (after the PERMATApintarTM Center, National University of Malaysia) for talented students aged 13 to 17 at the University of Sains Islam Malaysia (USIM) is PERMATA Insan College. The primary aim of this program is to cultivate and grow highly creative and talented Muslim children with a high degree of intelligence quotient (IQ) capable of becoming tomorrow's innovators, designers and presidents, scientists, researchers, or practitioners, yet imbued with the teachings of the Quran and Sunnah. PERMATA Insan College's location at USIM is in line with the university's ideology of combining educational knowledge of Naqli (revealed) and Aqli (acquired) [16].

Moreover, with the introduction of Naqli and Aqli awareness, PERMATA Insan 's instructional strategy is applied through the Gifted Curriculum program. The talent and talent of PERMATA Insan students are polished by tafseer (Quranic and Hadith perspective) by learning of science, technology, engineering, and mathematics subjects using ICT, memorization of the Quran and deepening understandings by Hadith studies and the application of DMGT in the implementation of the curriculum. PERMATA Insan College's comprehensive program system can educate skilled and talented students to be systematic, competent, and balanced persons with noble human personalities as the pillars for the creation of excellent generations and educated societies [16-20].

Intelligence may, in general, be defined as cognitive ability obtained by perception, listening and experience through the perception of learning. Quotient is another quantification notion that was originally applied to quantify the extent of the intellect of a person. When a person can consider something more complex, he may be rated as possessing a high degree of intellect. Knowledge is most sometimes referred to as thinking skills, problem-solving, multilingual, and a fast learner [24].

Authors in [20] proposed the concept of intelligence as the general capacity of an individual to act and respond efficiently in coping with the world in a purposeful, logical way. The most critical aspect of describing intelligence on behalf of Sternberg [21] is the actions exhibited. He thus articulated the term intellect as an adaptive action contributing to the achievement of a clear vision. Gardner [22], on the other hand, pointed out knowledge from the point of view of the individual's capacity to reason directly. He then suggested information as a capacity to address real issues. Gardner also indicated that people with intelligence were also able to identify or generate challenges that contributed to new knowledge being found. The quotient of intelligence or best known as IQ corresponds to the score in the form of statistical formulas calculated by dividing the chronological age with the mental age [23]. This will multiply the data obtained by 100. The average IQ rating is about a hundred. Scores achieved above the average of 100 are rated as high and can be considered low below the average of 100 [24-28].

### **3. Proposed Methodology**

#### **A. Situation analysis**

There are three main parts in the talented education. The awareness of the talented students is the first part, then the identification of these students forms the second part, finally the third part is based on offering and implementation of a talented program. This offered program should be based on differentiated and supportive curriculum.

Research studies presented two categories of approaches for identifying talented students namely, implicit and explicit approaches. The implicit approach investigates the implicit things in students then analyzes these things to identify the characteristics of talent that traditional IQ tests do not measure. Although the characteristics of talent are explained in a simple and easy content, its low robustness forms a barrier towards its application for real cases. On the other hand, the explicit approach depends on collecting large amount of data from students to be analyzed. The data collected in the form of answers to a prepared set of questions. However, the immature intellectual ability of students might now allow them to pass this list of questions.

According to Shaqra university's regulations, students must complete certain number of credit hours to be graduated. These credit hours are distributed over a set of years and semesters. In addition, students in each semester can select their courses and plan for their semester schedule. The identification of talented students can help them to speed up their graduation process. In addition, a new curriculum can be offered to these students to boot their skills. Moreover, these talented students can improve their talent by participating in regional and international competitions that require their talented skills.

#### **B. System Architecture**

The overall architecture of the proposed approach is depicted in Figure 1. In this figure, the process starts with collecting the academic records of the university students. The next step is filtering these records and extracting the points achieved by each student in a set of subjects that require certain type of talent. The points of these subjects are used to measure a set of features for further analysis. The extracted features are used to train a machine learning model to identify the students that are talented based on the predefined set of talent types. Once talented students are identified, we can make a special learning guide to them based on a talent quotient measured from the extracted features using neural networks. On the other hand, we can categorize these talented students based on their common characteristics. There categorization can help us to focus on certain characteristic and improve it at these students by encouraging them to participate in local or international competitions that can maximize their skilled characteristics.

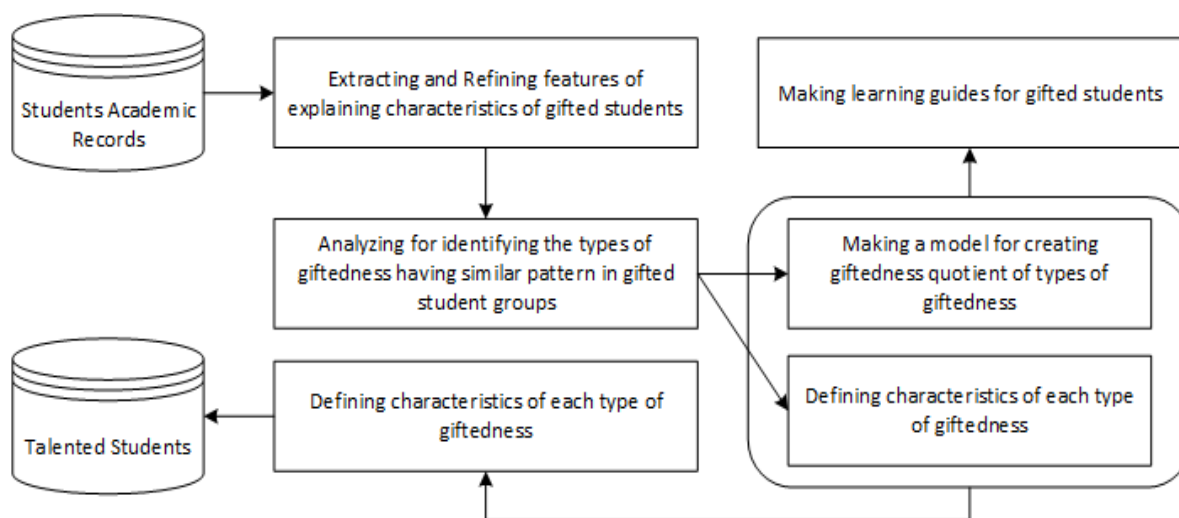


Figure 1: The process of identifying talented students

**C. Talented skills and Features**

As the proposed methodology targets identifying the talent of students of computer science major, this methodology is based on measuring the academic performance of students to identify the following list of talented skills:

1. Problem solving
2. Creativity
3. Memorization
4. Analysis
5. Communication skills

Each course in the study plan of computer science students partially contributes in measuring these skills. Therefore, we employed the academic records of students to evaluate the percentage of these skills based on their grades. For simplicity, the courses that student has already failed are excluded from the evaluation of these talented skills.

Table 1 shows the percentage of contribution of sample courses in identifying these talented skills. As shown in this table, depending on the nature of each course, the percentage of its contribution in determining the talented skills is predefined. For example, the programming courses depend mainly on problem solving and analysis, therefore their percentage for these two skills is high. In addition, it can be noted from Table 1 that the total score for each course is 100 and this score is distributed over the five skills of interest depending on the topics taught in these courses.

Table1: Percentage of talented skills for sample courses

Course	Talent				
	Problem solving	Creativity	Memorization	Analysis	Comm. skills
Programming I	50	20	5	15	10
Programming II	50	20	5	15	10
Statistics	65	-	5	15	15
Digital design	40	20	5	30	5
Discrete math	60	-	10	10	20
System analysis	15	-	5	65	15
Software engineering	10	10	10	60	10

For each skill denoted by  $S_j$ , we compute student's level denoted by  $L_j$  by considering all the courses that already completed. This is given by the following equation:

$$L_j = \frac{\sum_{i=1}^n G(C_i) \times P_j(C_i)}{\sum_{i=1}^n P_j(C_i)} \tag{1}$$

where,  $G(C_i)$  denotes student grade in course  $C_i$ ,  $P_j(C_i)$  denotes the percentage assigned to the skill  $j$  to the course  $C_i$ .

The output from this analysis phase is a feature vector for each course studied by the student. This feature vector consists of five values (each value denotes the score measure for one of the talented skills). The measured feature vectors are used in training a machine learning algorithm to classify student based on their talented skills. The training set is depicted in Figure 2, where each row represents the feature vector for a student  $S_m$ .

$S_1$	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$
$S_2$	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$
	...				
$S_M$	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$

Figure 2: The structure of feature vectors used in training machine-learning model.

**D. Clustering features**

In order to find the talented students, we apply the unsupervised machine learning algorithm called K-Means. This algorithm is used for clustering and classification. This algorithm is used to cluster the extracted feature vectors of in a group of  $K$  clusters based on the similarity among them. One the clusters are identified, we can define the characteristics of each cluster. Figure 3 shows the behavior of K-Means algorithm. In this figure, the data points in the left axis represents the feature vectors extracted for each student, where each point in this graph corresponds to a student. It can be noted that the extracted feature vector is composed of 5 dimensions. Therefore, for simplicity, we show only the first two or three dimensions. The right axis shows the result from the K-Means algorithm, in which the data points are partitioned into separate groups based on the similarity of its data points.

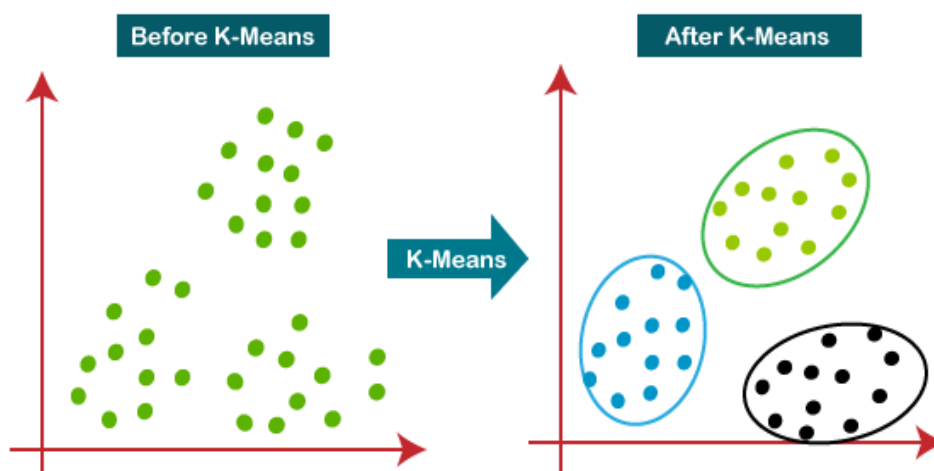


Figure 3: K-means clustering algorithm.

The steps of the K-Means algorithm are shown in Figure 4. The first step in this figure is the random selection of  $K$  centroids for the data points. Then, after repeating the steps of expectation and maximization, the final output will be as shown in Figure 3.

---

#### Algorithm 1 $k$ -means algorithm

---

- 1: Specify the number  $k$  of clusters to assign.
  - 2: Randomly initialize  $k$  centroids.
  - 3: **repeat**
  - 4:   **expectation:** Assign each point to its closest centroid.
  - 5:   **maximization:** Compute the new centroid (mean) of each cluster.
  - 6: **until** The centroid positions do not change.
- 

Figure 4: Steps of the K-means clustering algorithm.

The resulting clusters from K-means algorithm are used to categorize students into a set of types (from type A to type C). Figure 5 shown the results from applying k-means on our real dataset. The number of clusters is 3. Therefore, we now have three types of clusters from A to C, where each type consists of students with similar type of talent.

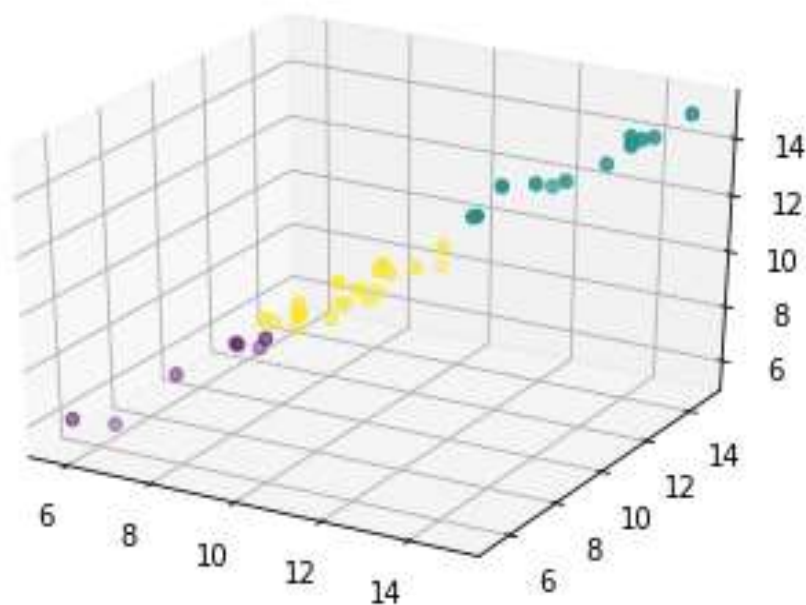


Figure 5: Clusters resulted from k-means on our dataset. For simplicity, we draw only the first two dimensions.

The evaluation of the value of  $K$  for the K-means algorithm is based on evaluating the best suitable number of types of talent that has low inter-cluster similarity and high intra-cluster similarity. Therefore, for our case we determined the value of  $k$  as 3 due to the similarity in the values in the extracted feature vectors.

#### E. Talent quotient

The similarity between the characteristics of students can be measured by the criterion called talent quotient. This criterion can be used to determine the type of talent based the extracted features. Therefore, talented students are distinguished by high value of talent quotient. The measurement of this talent quotient is achieved by the means of neural networks (NNs). There are two advantages in using NNs. Firstly, it is simple to determine the type of talent (type A, B, or C) using NNs. Secondly, the significance degree of the extracted features in determining the type of talent can also be measured based on it.

$L1$	$L2$	.....	$L_n$	Cluster	Output
5	3		7	C1	100
3	2		6	C2	1
3	2		4	C3	1

Figure 6: Preparation of data used in training multi-layer neural network.

To feed the NN with the input data, we first identify the classes clustered by the k-means algorithm and use these classes to find determine the input and output from NN. Figure 6 shows the data prepared for the result of the k-means algorithm. In this figure, for each feature vector, we determine the class to which it belongs. In addition, for the best type of clusters we put a value 100 as an output from the neural network. Therefore, the input to NN becomes the feature vectors representing the measure of the skills, and the output is a value from 1 to 100 representing the talent quotient.

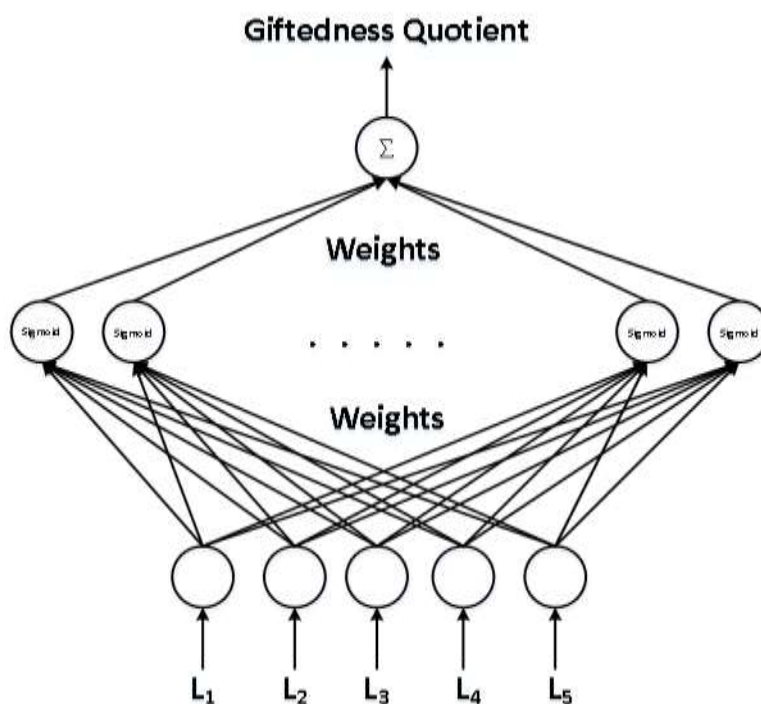


Figure 7: Prediction of talent quotient based on multi-layer neural network.

The multi-layer neural network used in this work is depicted in Figure 7. This neural network is composed of a set of neurons forming the input layer, another set of neurons forming the middle/hidden layer, and one neuron in the output layer. The number of input neurons equal to the length of the feature vector, which is 5 in our case. The number of neurons in the hidden layer usually determined as double or triple the number of neurons in the input layer. In our model, we used 25 neurons in the hidden layer. Moreover, the activation function used for the hidden and output neurons is the sigmoid function. To measure the talent quotient, the back-propagation algorithm is employed.

To create the talent quotient, the following procedure is used for building the neural network.

1. A part of the survey data is randomly selected and a target output value of 100 is assigned to the cluster that belongs to is Type X, and value of 1 is assigned to other types.

2. To reduce the difference between the output value from the neural network and the actual target output, back-propagation algorithm is employed in iterations until a termination criterion is satisfied.
3. After a set of iterations and at the end of the training process, the neural network model for creating the talent quotient becomes ready.
4. When testing the neural network model, we decide that the student belongs to type C of talent if the output from the neural network is close 100.

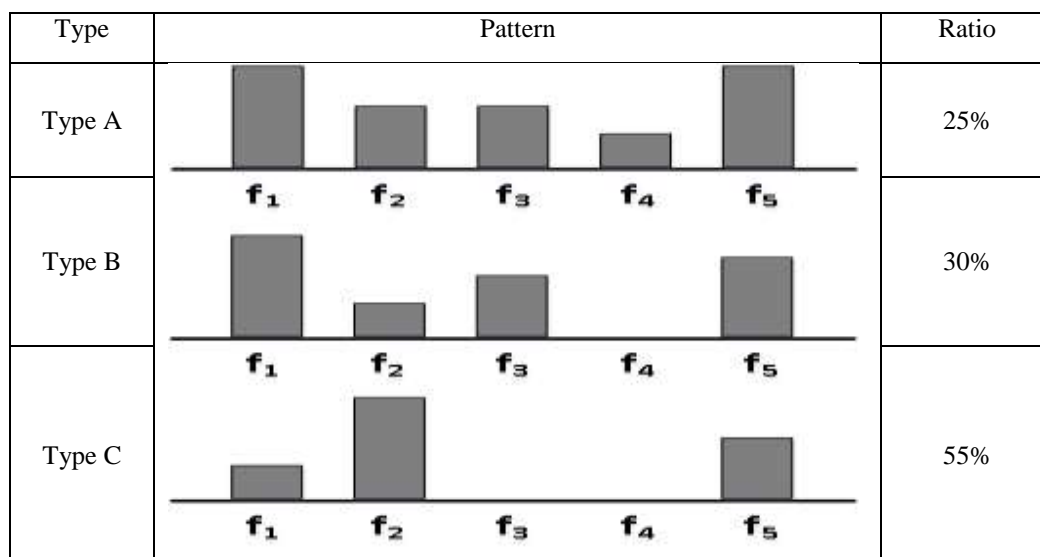


Figure 8: Characteristics of talent types, where f1 to f5 denote the skills measured from students academic records.

Based on the collected set of student records, we discovered three talent types having similar features. Figure 8 depicts these types and the ratio between the mean number of students in each type with respect to the total number of students. As shown in the figure, type A consists of the most talented students, where the ratio of them is 25%. These types consist of students that are talented with the skills: f1 (Problem solving), f2 (Creativity), f3 (Memorization), f4 (Analysis), and f5 (Communication skills).

As shown in Figure 8, it can be noted that students of type A, for example, have high value of f1 and f5 that refer to problem solving and communication skills. Therefore, we can encourage these students to gain more experience and improve these skills through getting them participating in local and international competitions that are based on these skills, so that maximizing both problem solving and communication skills. In addition, during their study, we can stimulate their attitude to improve these skills with them.

#### 4. Conclusion

In this paper, a machine learning based approach is proposed for identifying the talented students from their academic records. The implicit capabilities of talent students are measured based on their performance in the university study and then classified into a set of types of talent. To do this classification, k-means clustering is applied to find the various types of talent along with their characteristics. To evaluate the similarity between characteristics of students and types of talent, a multi-layer neural network was used to predict the talented quotient..

**Funding:** “This research received no external funding”

**Conflicts of Interest:** “The authors declare no conflict of interest.”

**References**

- [1] Imison, K., *The Acceptance of Difference*, Report on the review of gifted and talented education in Queensland's state school, Education Queensland, 2001..
- [2] Renzulli, J.S. & Smith, L., Two approaches to identification of gifted students. *Exceptional Children*, 43(8), pp. 512–518, 1977.
- [3] Education Queensland Report, *A Model for Curriculum Provisions for Gifted Education and Talent Development*, Education Queensland, 2000.
- [4] Department of Education and Arts, *Framework for Gifted Education*, Queensland Government, 2004.
- [5] Renzulli, J.S., The identification and development of giftedness as a paradigm for school reform. *Journal of Science Education and Technology*, 9(2), pp. 95–114, 2000.
- [6] Gardner, H., *Intelligence Reframed: Multiple Intelligences for the 21st Century*, Basic Books: New York, 2000.
- [7] Clark, B., *Growing Up Gifted: Developing the Potential of Children at Home and at School*, Prentice Hall: New York, 2001.
- [8] Quinlan, J.R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann: San Mateo, CA, 1993.
- [9] Chen, M.S., Park, J.S. & Yu, P.S., Data mining for path traversal patterns in a web environment. *Proc. of the 16th Int. Conf. on Distributed Computing Systems*, 1996.
- [10] Kosala, R. & Blockeel, H., Web mining research: a survey. *SIGKDD Explorations*, 2(1), pp. 1–15, 2000.
- [11] Almeida, L.S., Prieto, L.P., Ferrando, M., Oliveira, E. & Ferrándiz, C. (2008). Torrance test of creative thinking: The question of its construct validity. *Thinking Skills and Creativity*, 3,1, 53-58.
- [12] Bélanger, J. & Gagné, F. (2006) Estimating the size of the gifted/talented population from multiple identification criteria. *Journal for the Education of the Gifted*, 30(2), 131-163.
- [13] Brown, S.W., Renzulli, J.S., Gubbins, E.J., Siegle, D., Zang, W. & Chen, C-H. (2005). Assumptions underlying the identification of gifted and talented students. *Gifted Child Quarterly*, 49 (1), 68-79.
- [14] Callahan, C. M., Tomlinson, C. A., Hunsaker, S. L., Bland, L. C., & Moon, T. (1995). *Instruments and evaluation design used in gifted programs*. (RM 95132). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut
- [15] Davis, G.S, Rimm, S.B., & Siegle, D. (2011). *Education of the gifted and talented* (6 th Ed). New York: Pearson. Educational Planning and Research Division. (2008). *Education in Malaysia: a journey to Excellence*. Putrajaya: Ministry of Education.
- [16] Gagné, F. (2010). Motivation within the DMGT 2.0 framework. *High Ability Studies*, 21(2), 81-99
- [17] Goodhew, G. (2009). *Meeting the needs of gifted and talented students*. London: Continuum International Publishing Group.
- [18] Han, K. S., & Marvin, C. (2000). A five year follow-up study of the Nebraska Project. *Roeper Review*, 23 (1) ,25-33.
- [19] McBee, M. (2010). Examining the probability of identification for gifted programs for students in Georgia Elementary Schools: A multilevel path analysis study. *Gifted Child Quarterly*, 54, 283-297.
- [20] Noriah, M. I., Rosadah, A.M., & Siti Fatimah, M.Y. (2009). *PERMATApintar: Pengalaman UKM*, (In Malay), Bangi: Universiti Kebangsaan Malaysia
- [21] Peters, S.J., & Gentry, M (2010). Multigroup construct validity evidence of the HOPE scale: Instrumentation to identify low-income elementary students for gifted programs. *Gifted Child Quarterly*, 54, 298-313.
- [22] Piirto, J. (1999). *Talented children and adults; their development and education*. 2nd. Ed. Ohio: Merrill Prentice Hall
- [24] Purcell, J. H., & Eckert, R.D. (2006). *Designing services and programs for high-ability learners*. Thousand Oaks: Corwin Press.

- [25]Putallaz, M., Baldwin, J. & Selph, H. (2005). The Center for Talented Youth talent search and academic programs. *High Ability Studies*, 16, 41-54Siti Fatimah Mohd Yassin et al. / *Procedia - Social and Behavioral Sciences* 55 ( 2012 ) 585 – 593
- [26]Renzulli, J.S. (2002). A practical system for identifying gifted and talented students. In Renzulli, J. S., Smith, L. H., White, A. l.,
- [27]Callahan, C. M., Hartman, R. K.,& Westberg, K. L. Scales for rating the behavioral characteristics of superior students: Revised edition. Technical and Administration Manual. (pp 46-54). Mansfield Center: Creative Learning Press.
- [28]Rigby, K. (2005). ‘Rocky Mountain Talent Search’ at the University of Denver. *High Ability Studies*, 16 (1), 71-75.