Implementation K-nearest neighbour for student expertise recommendation system

I Taufik*, Y A Gerhana, A I Ramdani and M Irfan
Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia

*ichsan@uinsgd.ac.id

Abstract. The ability of students to determine their chosen field of expertise is still subjective, many students choose the field of expertise because their classmates choose the field of expertise not by considering their abilities and interests. This research uses the KNN classification method to determine areas of expertise that are in accordance with student expertise. The KNN method was chosen because it is a method that uses supervised algorithms where the results of new query instances are classified based on the majority of the categories in the KNN whose purpose is to classify test data based on training data. This system was tested using the confusion matrix method and the results were 98.30% of the total student data sample of 30 people.

1. Introduction
Higher education is one of the most important educational institutions to produce professional staff. At the tertiary level a person learns based on a particular field of expertise, so that the output of a college is expected to be able to produce quality human resources in accordance with their field of expertise [1]. Based on this, it is very important for a student to study in accordance with the field of expertise and focus on that field.

However, many students choose areas of expertise with subjective considerations such as just because classmates choose the area of expertise [2]. The problem of selecting a field of expertise arises when the abilities possessed by students are not the same as their interests [3]. To deal with these problems, a system will be built that is able to provide advice on areas of expertise in accordance with student expertise. This system is a computer-based information system that approaches to produce various alternative decisions to assist certain parties in handling problems using data and models [2].

One of the supporting methods suitable for classification is K-Nearest Neighbour (KNN) because this method is a method that uses supervised algorithms where the results of new query instances are classified based on the majority of categories in the KNN whose purpose is to classify test data based on data practice [4].

2. Methodology

2.1. KNN method
K-Nearest Neighbour is a classification algorithm based on analogy, which compares testing data with training data that is close and has similarities with the training data [5]. In Figure 1 is the K-Nearest Neighbour Algorithm flowchart.
2.2. Software development method
The prototype model is one method of software development that is widely used by developers because in addition to being simple, this method can also adjust user requirements in detail in making software. This method is able to offer the best approach in terms of certainty about the efficiency of the algorithm, the ability to adjust from an operating site or the forms that must be done by human-machine interaction [6]. The cycle of this prototype model can be seen in Figure 2:

**Figure 1.** KNN algorithm flowchart.

**Figure 2.** Prototype paradigm [6].
3. Result and discussion
Criteria and weighting values can be seen in Table 1.

| Criteria             | Weight | Weight (%) |
|----------------------|--------|------------|
| GPA                  | 5      | 25         |
| Interest             | 8      | 40         |
| Elective Courses     | 7      | 35         |

Case example if:

Name : Budi
Areas of expertise that are of interest : Computer Systems and Distributed Computing
Subjects of interest : Computer organization & architecture, computer networks, distributed systems
GPA semester 1 - 6 : 3

The manual calculation of the proximity of training data with testing data is in accordance with the proximity formula as follows: in this example the calculation of the case with the first sample.

Proximity to the 1st sample = \[ \sum \frac{(nx*bnx)}{(bnx)} = 0.44 \]

This closeness calculation is repeated as many as 10 existing sample data. And the results of calculations with 10 sample data can be seen in Table 2.

| No | Proximity | Result \((H1+H2+H3)\sum bn\) |
|----|-----------|-------------------------------|
| 1  | Sample 1  | 0.44                          |
| 2  | Sample 2  | 0.875                         |
| 3  | Sample 3  | 0.565                         |
| 4  | Sample 4  | 0.51125                       |
| 5  | Sample 5  | 0.75                          |
| 6  | Sample 6  | 0.675                         |
| 7  | Sample 7  | 0.625                         |
| 8  | Sample 8  | 0.825                         |
| 9  | Sample 9  | 0.74                          |
| 10 | Sample 10 | 0.8                           |

After calculating the closeness as much as 10 times, then look for data with the highest closeness value. In this study using a range of proximity 0 to 1. Then the results taken are sample data that has the highest closeness value. From Table 2, it can be seen that the area of expertise suggested in the case example is the field of expertise that corresponds to the sample 2 with a value of proximity of 0.875.

4. Implementation
In Figure 3 is the implementation of the system interface for the KNN classification page.
4th Annual Applied Science and Engineering Conference  
Journal of Physics: Conference Series 1402 (2019) 077004  
do:10.1088/1742-6596/1402/7/077004

Figure 3. Classification.

On this page students enter the data needed for the KNN classification calculation process which will then be carried out the calculation process and the results will look like in Figure 4.

Figure 4. Classification result.

5. Conclusion
The conclusions that can be drawn from this study are:

- The KNN algorithm can be applied to the application of recommendations in the field of student expertise by applying the calculation of the KNN classification.
- Based on the results of testing 30 samples using confusion matrix, the KNN algorithm on this system has an accuracy of around 98.30%.

Acknowledgments
Authors wishing to acknowledge Research and Publication Centre of UIN Sunan Gunung Djati Bandung that supports and funds this research publication.

References
[1] Niswatin R K 2013 Penempatan Jurusan Mahasiswa Baru Menggunakan Metode K-Nearest
Cogito Smart Journal 11 55–67

[2] Muhammad F L Keahlian Menggunakan Metode Weighted Product (Studi Kasus: Jurusan Teknik Informatika)

[3] Ridwan M, Irfan M and Taufik I 2017 Fitur Rekomendasi Mata Kuliah Pilihan dengan Menggunakan Algoritma Fuzzy Analytical pada Aplikasi KRS Berbasis Mobile J. INSIGHT 11

[4] Taufik I, Syaripudin U and Jumadi J 2017 Implementasi Metode Promethee Untuk Menentukan Penerima Beasiswa J. Istek 10 1 21

[5] Salamun S 2018 Penerapan Algoritma Nearest Neighbor dan CBR pada Expert System Penyimpangan Perilaku Seksual J. Online Inform. 22 63

[6] Pressman R S 2012 Rekayasa Perangkat Lunak (Yogyakarta: Andi)