JOIN (Jurnal Online Informatika)
Volume 6 No.1 | June 2021: 53-62
DOI: 10.15575/join.v6i1.701
p-ISSN: 2528-1682
e-ISSN: 2527-9165

Prediction System for Problem Students using k-Nearest Neighbor and Strength and Difficulties Questionnaire

Dede Kurniadi[1], Asri Mulyani[2], Inda Muliana[3]
[1,2,3] Department of Informatics, Sekolah Tinggi Teknologi Garut, Indonesia

Article Info

Article history:

Received February 10, 2021
Revised April 05, 2021
Accepted April 15, 2021
Published June 15, 2021

Keywords:

Development
K-Nearest Neighbor
Prediction System
Problem students
Questionnaire
Rapid Application
Strength and Difficulties

ABSTRACT

The student counseling process is the spearhead of character development proclaimed by the government through education regulation number 20 of 2018 concerning strengthening character education. Counseling at the secondary school level carries out to attend to these problems that might resolve with a decision support system. So that makes research challenging to measure completion on target because it is not doing based on data. The counseling teacher does not know about student's mental and emotional health conditions, so it is often wrong to handle them. Therefore, we need a system that can recognize conditions and provide recommendations for managing problems and predicting students who have potential issues. The Algorithm used to predict problem students is K-Nearest Neighbor with a dataset of 100 students. The stages of predictive calculation are data collection, data cleaning, simulation, and accuracy evaluation. Meanwhile, building the system is done using the rapid application development methodology where the instrument used to map the student's condition is the Strenght and Difficulties Questionaire instrument. This research is a system to predict problem students with an accuracy rate of 83%. The level of user experience based on the User Experience Questionnaire (UEQ) results in the conclusion that the system reaches "Above Average.". This system is expecting to help counseling teachers implement an early warning system, help students know learning modalities, and help parents recognize the child's personality better.

Corresponding Author:

Dede Kurniadi,
Department of Informatics,
Sekolah Tinggi Teknologi Garut,
Jl. Mayor Syamsu No. 1 Jayaraga Garut 44151 Indonesia
Email: dede.kurniadi@sttgarut.ac.id

1. INTRODUCTION

A study initiated by the World Health Organization and the World Bank is called The Global Burden of Disease [1]. It predicts that mental illness will rank second after cardiovascular disease in 2020, which will cause social and economic burdens. Meanwhile, in Indonesia, mental disorders have not been seen as a severe threat. Based on the 2018 RisKesDas data, it was founding that [2] the prevalence of mental and emotional disorders for those aged 15 years and over was in the range of 6.1% of the total population of Indonesia, where at the age of 15, the Indonesian population was still a student in middle school, this would undoubtedly be very important. Influence on the educational process.

To overcome this, the government, through Law No. 20 of 2003, has made the counseling program in Educational units the spearhead for overcoming student's mental and emotional problems. For the counseling process to be effective, it is necessary to identify students who need special attention regarding mental and emotional disorders. One of the instruments that can determine a student's mental condition is The Strengths and Difficulties Questionnaire (SDQ) [3], an instrument capable of mapping the mental and emotional health conditions of children aged 4 to 17 years.

The SDQ instrument consists of 25 questions divided into five categories, namely emotional symptoms (E), behavioral problems (C), hyperactivity (H), peer problems (P), and prosocial strength (Pr), where the results of USQ are to classify students. Into three categories, namely normal, borderline and abnormal. With

the limited resources of educators in the secondary school environment, we need a system that can predict potentially problematic students based on USQ indicators. So that it is necessary for an early warning system so that counseling teachers can focus attention on the right students.

A system to help Counseling teachers identify and predict student problems more quickly, the solution is to build a prediction system that can classify student problems in the future by using USQ results data as a dataset in predicting student problems. One of the classification methods used to classify problematic students is the K-Nearest Neighbor (k-NN) algorithm, which can rank based on the similarity to the previous data. The general formula for finding distances using the Euclidian [4].

There are several previous studies related to this study. The first [5] examines the effect of learning styles, family environment, and facilities on achievement; the results show that these components positively affect student achievement. The second study [6] examined gender, age, type of residence, the number of Semester Credit Units (SKS) to predict Quality Score (NM) using the k-NN Algorithm. The third study [7] examines data mining applications to predict student achievement based on socioeconomic, motivation, discipline, and past achievements using several methods, namely the J48 decision tree algorithm, CHAID, and multiple regression analysis. The fourth study [8] examined how to predict vocational students' behavior using the Naive Bayes method. The fifth study [9] indicated student graduation. The sixth study [4] discusses the prediction of scholarship acceptance in tertiary institutions with 4 (four) variables, namely semester, GPA, parents income, dependents, with an accuracy of 95.83%.

Based on the previous paragraph's research, it is necessary to research the k-Nearest Neighbor algorithm's implementation on the prediction system for problem students using rapid application development. Where the system can identify mental conditions, emotional conditions, provide treatment recommendations, and provide predictions for students who are considered potentially problematic.

2. METHOD

2.1. Research Framework

In this research, the system development uses a rapid application development (RAD) approach, which allows the system to be built in a short and fast time [10]. RAD is a process model that can construct linear sequential software that emphasizes a concise development cycle with an estimated 60 to 90 days [11]. RAD has four stages, namely business modeling, data modeling, process modeling, and application building. In contrast, in the Build Application stage, there are data mining or knowledge discovery in database (KDD) stages or those that function to extract information, in this case, in the form of prediction based on data. [12], where the data obtained comes from respondents who filled out the Strengths and Difficulties Questionnaire instrument.

In this study, four activities were carried out to predict the outcome [4], namely data collection, data processing, K-NN calculation, Modeling Simulation, and accuracy testing. Then, to measure the system's user experience is built using the User Experience Questionnaire (UEQ). This instrument can use to determine the extent of user experience using a program [13]. The research framework, which includes RAD activities as system development and KDD as a reference for research activities, is described in the research framework in Figure 1 :

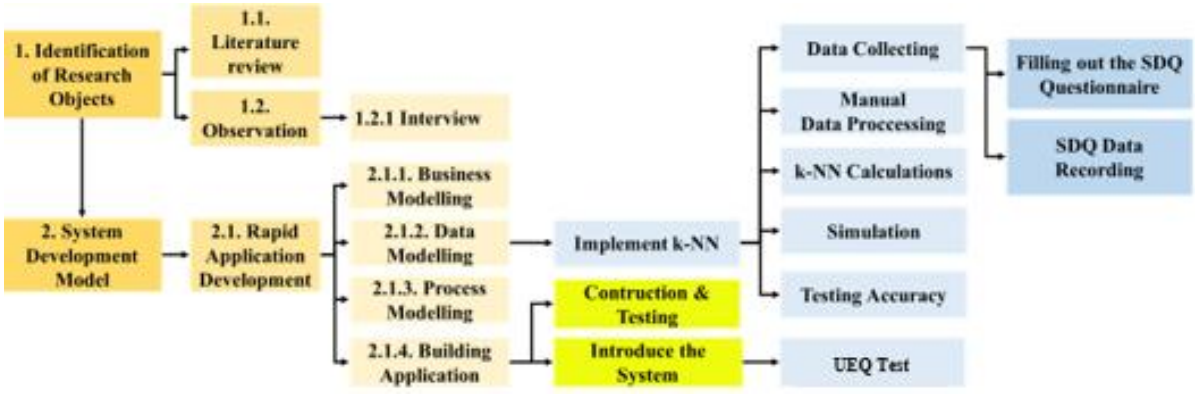

Figure 1. Research Framework

A brief explanation based on Figure 1 regarding the research framework, carried out with 2 (two) main activities, namely:

1. Identification of research objects is collecting data on the thing under study by conducting literature studies and interviews with counseling teachers at Wikrama 1 Garut Vocational High School, namely Mr. Muhamad Nurjalil and the counseling coordinator of Tarogong Kaler Health Center.
2. The next activity is system development with the RAD method approach, in which there is the activity of implementing the k-Nearest Neighbor algorithm. The system's results serve as a form for students to fill out the SDQ and perform calculations. The data from the calculation results will then be used as predictive data to conclude that the student includes potentially problematic or not a category.

2.2. k-Nearest Neighbor

K-Nearest Neighbor is an algorithm that performs predictions by classifying it to designate something based on similarities to previous data [12]. The formula commonly used to determine distances using Euclidean is as in equation (1).

$$d_i = \sqrt{\sum_{i=1}^{p}(x_{2i} - x_{1i})^2} \quad (1)$$

Information : d : distance; p : data dimension; i : variable; x_{2i} : sample data; x_{1i} : testing data.
In general, the k-NN algorithm process is as follows.

1. Preparing sample data in the form of an array;
2. Preparing testing data in the form of an array;
3. Calculating the distance between attributive values of testing to each training using Euclidean Distance;
4. Sorting the distance results based on the lowest values and the predetermined number of neighbors;
5. Obtaining the prediction results based on the calculation of the highest number;
6. Calculating the accuracy based on the prediction;

The use of the k-NN Algorithm as a prediction system algorithm is due to the characteristics of the data set. Where five attributes are numeric, and one target data is nominal. The data features are based on data mining roles theory [14], which can use as a classification that can accommodate numeric and nominal attributes, but the target must be in nominal form.

2.3. Strengths and Difficulties Questionnaire

Strengths and Difficulties Questionnaire or SDQ is a parameter used in psychology to map the strengths and weaknesses of students who need more attention, which usually have a relationship with emotional and behavioral. Using SDQ, educators can determine the child's actual condition according to the data to find out children with special needs [15], which defines child behavior that deviates from the majority of normal children in terms of mental, sensory, physical, and neuromuscular abilities. SDQ [3] consists of 25 questions with five dimensions that can measure prosocial, hyperactivity, emotional problems, and behavior towards peers.

.

2.4. User Experience Questionnaire

User Experience Questionnaire Is an instrument that can determine the extent of user experience using a program, namely the System Usability Scale (SUS), which uses UEQ to give an easy and efficient questionnaire to measure user experience. UEQ has 6 (six) assessment components, namely:

1. Attractiveness: whether the user likes the program subjectively
2. Perspicuity: easy or not to use
3. Efficiency: how simple the program building
4. Dependability: users feel in control of their interactions with the program
5. Stimulation: motivate users to use the program
6. Novelty: how creative and innovative.

The assessment components are broken down into 26 (twenty-six) questions, where each question has an answer scale from 1 - 7. EUQ uses English, but there have been several studies making UEQ in the Indonesian version. The data obtained is entered into a UEQ Data Analysis Tool tool with the .xlsx format that previous researchers have provided. [13] the output of the data has been automatically generated by the tool visually.

2.5. Counseling

Counseling can not be separate from the word guidance, where counseling [16] discusses individual problems in everyday life, both personal and general issues. Rochman and M Surya (Natawijaya) also argued that Counseling is a form of relationship between two people. One of them acts as a client who is assisting in adjusting themselves to their environment effectively. Based on this description, it can be concluded that Counseling is assistance provided to individuals to solve life problems by interview or by adjusting to their environment [17].

Guidance and Counseling is a systematic, objective, logical, ongoing, and programmed effort carried out by guidance and counseling teachers or counselors to facilitate student's/counselors development in achieving independence. Guidance and Counseling are integral components of the education system in every Educational unit, seeking to encourage and empower students/counselees to attain complete and optimal development. As an essential component, guidance and Counseling, which is independent in an integrated manner, synergizes with the administrative and management service areas and the curriculum and educational learning areas.

2.6. Testing Accuracy

To measure the accuracy model using the Confusion Matrix tool. It is commonly used to evaluate the classification model to predict correct and incorrect objects. In other words, it usually contains information on actual values and prediction on classification. What follows is the calculation formula of accuracy rate.

$$Accuracy = \frac{Number\ of\ True\ Value}{Total\ Data\ Amount} \times 100\% \quad (2)$$

3. RESULTS AND DISCUSSION

This research follows the stages in the research framework presented in section 2. The result is a system that can map student's personalities and make predictions for potential problems.

3.1. Object Identification

The research was conduct at SMK Wikrama 1 Garut, which was established in 2010 under the Prawitama Foundation auspices with the national school number (NPSN) 20275997. Every year a new student personality mapping is carried out in collaboration with the local Puskesmas.

3.2. Interview

In this study, an interview was conducted with a counseling teacher, and it was found that the instrument used to show the talent mapping was a Strengths and Difficulty Questionnaire [3], which produces learning modalities, namely student learning styles and brain dominance, and mental health, which includes emotional health, behavioral problems, hyperactivity, peer and prosocial problems. The mental health outcomes will then be used as predictive variables.

3.3. System Development

The next stage in RAD is to carry out business modeling, which aims to provide a visual description of the business model proposed in system design. The system development process is done by describing the business process, modeling the data, and modeling the system.

3.4. Business Modelling

The business process of the application that was built involved four parties, but only two actors were involved, namely, students and counseling guidance teachers where the Public health center only played a role in providing a questionnaire form originating from SDQ and receiving complex files from the compilation of student answers. The business model that occurs is described in Figure 2.

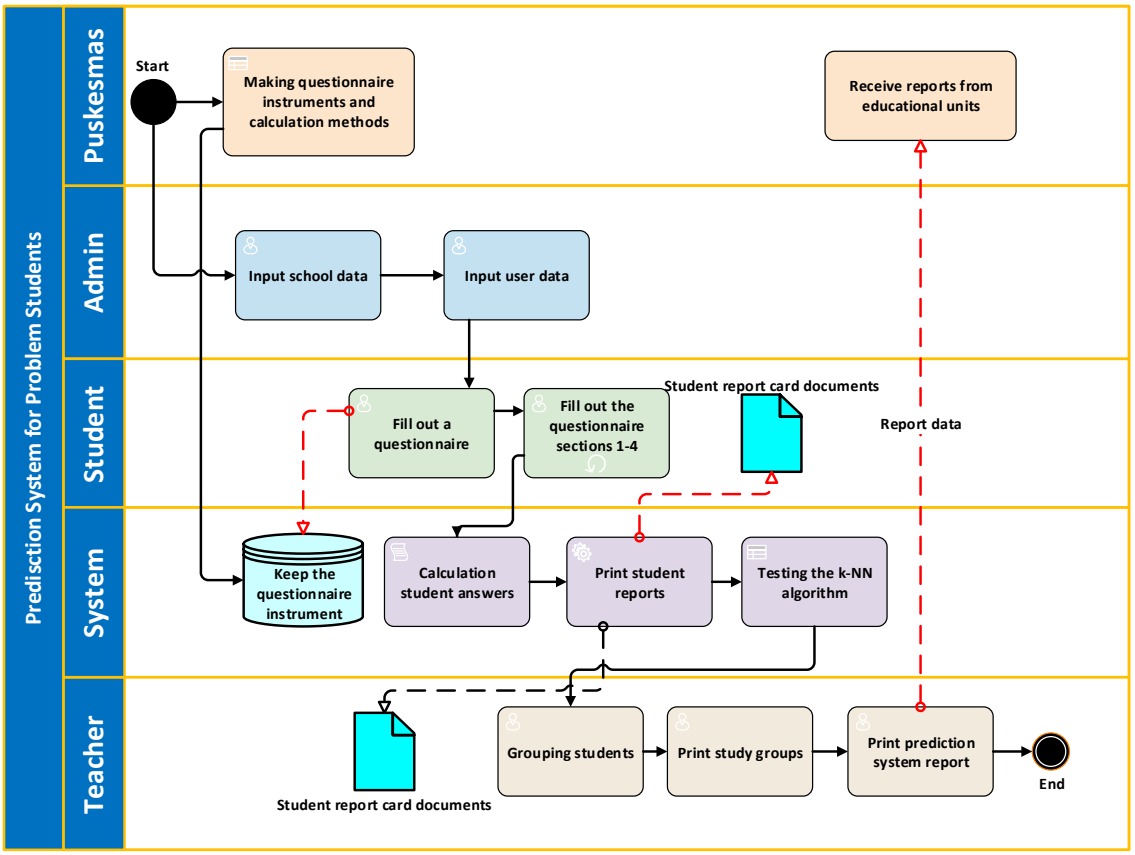

Figure 2. Business Modeling

3.5. Data Modelling

Application data modeling The depiction of databases in entity data relationships (ERD) is presented in Figure 3.

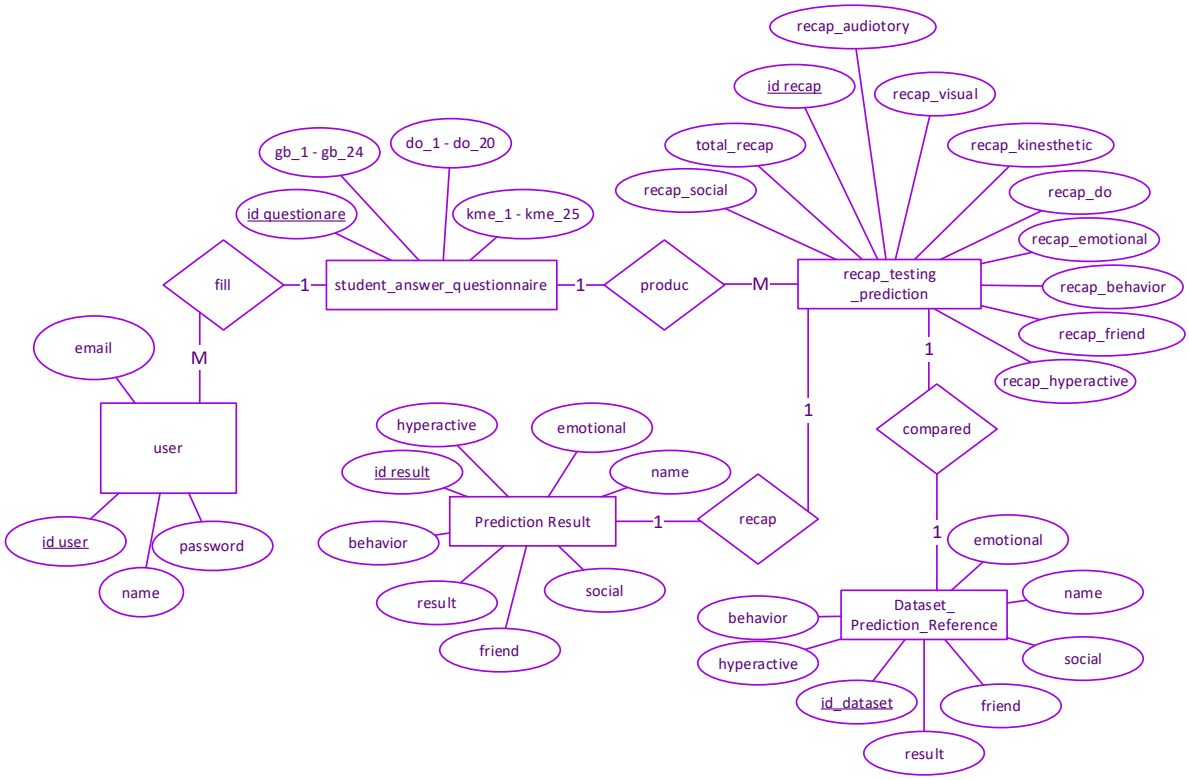

Figure 3. Entity Relationship Diagram

p-ISSN: 2528-1682
e-ISSN: 2527-9165

3.6. Process Modelling

Process modeling describes using a data flow diagram (DFD), which defines the interaction between data and actors interacting.

The Zero diagram plays a role in describing every process in the proposed zero system diagram. This zero diagram includes five processes, each of which consists of :

1) Input *user*, where the admin performs user input, both counseling teachers and students with different access rights for each user
2) Filling out the questionnaire, in this phase, the students fill out the questionnaire data where the questionnaire components come from the SDQ instrument, which is then storing in the database.
3) Calculation of report cards with SDQ calculations which produce five variables in the form of numerical data, which are then using as attributes in the calculation of the k-NN Algorithm
4) Prediction calculations have been finishing by comparing five attribute data and one target data in the dataset with test data.
5) Data grouping

The DFD of the system being building details in Figure 4.

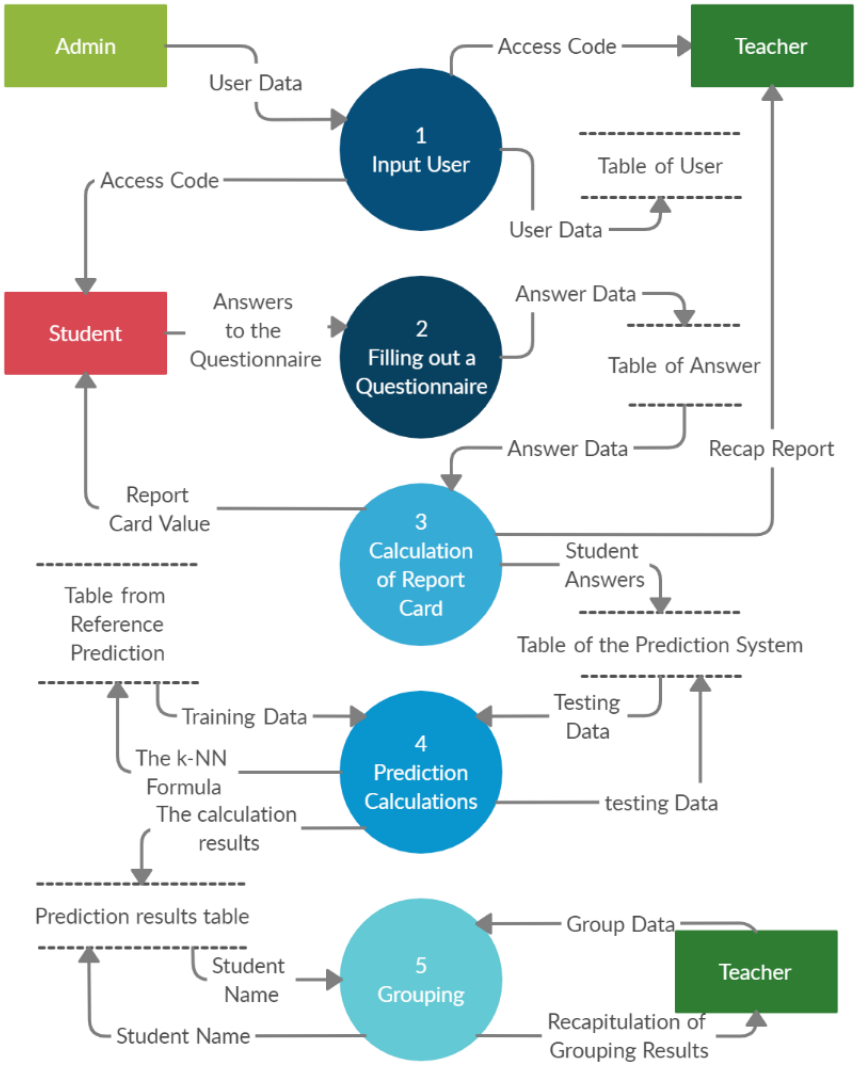

Figure 4. DFD level 0

3.7. Data Collecting

Data collection obtained data attributes in the form of student names, emotional data (E), behavioral problems (C), hyperactivity (H), peer problems (P), prosocial (Pr) personality report cards, and nominal target data with Good student grades or Not Very Good. Students are said to be not good if they have involved in a particular case.

3.8. Manual Data Processing

Manual data processing is the process of cleaning redundant, incomplete data. One hundred (100) student data had been clean and ready to be calculated, as shown in Table 1.

Table 1. Cleaning Result Data

No	Name	Personality					Result
		E	C	H	P	Pr	
1	Risa Herajaya	5	3	9	3	10	Good
2	Ichsan Muhammad B	4	3	6	4	7	Good
3	Muhamad Albi	7	3	8	6	9	Bad
..	…	..	..	..	..	..	…
100	Depi Tirawati	6	2	5	6	10	Good

Where E: emotional; C: behavior problems; H: hyperactivity; P: peer problems; Pr: prosocial

3.9. k-Nearest Neighbor Calculations

The process of calculating the prediction is carried out by taking five variables, namely E (emotional), C (behavioral problems), H (hyperactivity), P (peer problems), Pr (prosocial), to then look for results, which are categorized as good or bad. In principle, the calculation is done by looking for the most similar value from the dataset that has been obtaining. The detailed calculation process will be elaborated further in sub 3.13. simulation.

3.10. Simulation

The simulation process uses the K-Nearest Neighbor (kNN) algorithm by classifying it based on the similarity to the previous data [9]. The formula is commonly used to determine distance using euclidean [17]. The simulation process uses the K-Nearest Neighbor (KNN) algorithm with an analogy of student X is taken with the data shown in Table 2.

Table 2. Test Data

No	Name	E	C	H	P	Pr
1	Student X	2	4	4	3	8

The data is calculated using the euclidean formula into the dataset and obtained 5 (five) similar data to determine whether the student is classified as a Good or Bad student. The calculation output from the simulation is presenting in Table 4.

Table 4. Test Data Simulation Results

No	Name	E	C	H	P	Pr	D	R	K	Result
7	Fachrezi Nurdin Zaelani	2	4	5	4	7	1.7	1	Yes	Good
36	Muhammad Rafli	1	3	5	3	9	2.0	2	Yes	Good
40	Ramdan Abdul Malik	3	3	5	4	9	2.2	3	Yes	Good
53	Nela	2	2	4	3	7	2.2	4	Yes	Bad
57	Siti Aisah	3	4	6	3	9	2.4	5	Yes	Good

Based on Table 4, the data that has been sorting from the most minor 5 data were obtained, 4 of which are Good and one Bad. Students on behalf of Student X are declared good.

3.11. Testing Accuracy

Based on the testing data results, as many as 100 students tested on the data itself, 83 were accurate, and 16 were inaccurate. The accuracy result table is presenting in Table 5.

Table 5. Accuracy data

No	Name	E	C	H	P	Pr	Result	Prediction	Accuracy
1	Risa Herajaya	5	3	9	3	10	Good	Good	Be accurate
2	Ichsan Muhammad B	4	3	6	4	7	Good	Good	Be accurate
3	Muhamad Albi	7	3	8	6	9	Bad	Bad	Be accurate
4	Pooja Melaty S	4	1	8	3	8	Good	Good	Be accurate
5	Hermalia Musalimah	4	2	6	3	8	Good	Good	Be accurate
..	……	…	…	…	…	…	…	…	…

No	Name	E	C	H	P	Pr	Result	Prediction	Accuracy
100	Depi Tirawati	6	2	5	6	10	Good	Good	Be accurate

Based on the data that has been carrying out by experiments on 100 students against the data itself, 83 results are accurate, and 16 are inaccurate, so it can conclude that the percentage of accuracy obtained based on the calculation of formula (2) is 83%.

3.12. Building Application

This stage develops software for prediction systems using PHP programming with the Laravel framework and MySQL for data storage. This stage's result is a prediction system application. Here are some examples of how the page displays in Figure 5.

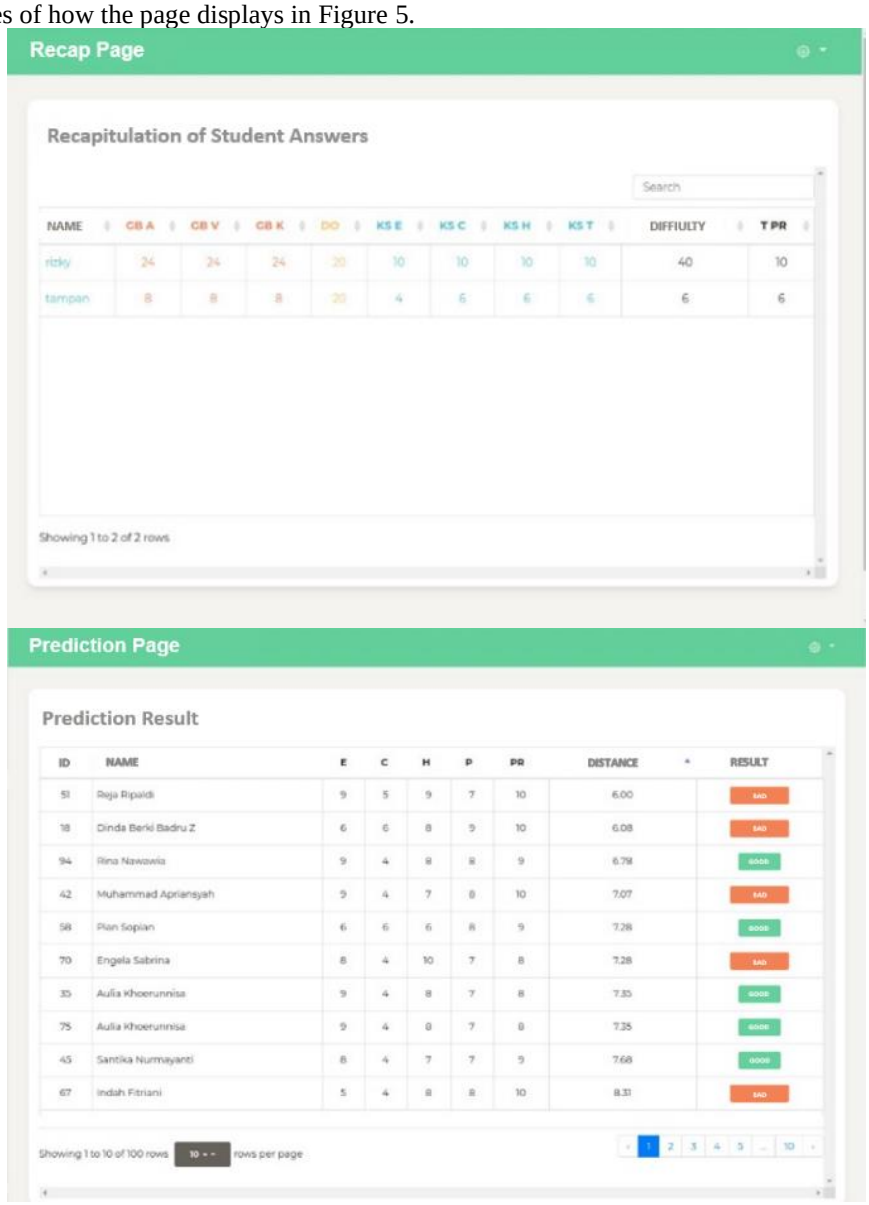

Figure 5. Example of the display system to predict problem students

3.13. Construction and Testing

At this stage, the researcher builds a program that has been planed. Program development includes meeting hardware and software requirements for software using the PHP language with the help of the Laravel framework.

3.14. User Experience Questionnaire Test

The system built is then introduced to the user by using a User Experience Questionnaire (UEQ), which can measure the user experience in using the system, consisting of 6 (six) components, namely attractiveness,

clarity, efficiency, accuracy, stimulation, and novelty. [13]. UEQ, which was a test on 10 (ten) people, 2 (two) counseling teachers, 2 (two) software engineers, and 6 (six) students, resulted in the conclusion that the system reached Above Average, as shown in Table 6 and Figure 6.

Table 6. Result User Experience Questionnaire Test

No	Scale	Mean	Comparison to benchmark	Interpretation
1	Attractiveness	1,52	Above average	25% of results better, 50 of results worse
2	Perspicuity	0,57	Bad	In the range of the 25% worst results
3	Efficiency	1,20	Above average	25% of results better, 50 of results worse
4	Dependability	1,18	Above average	25% of results better, 50 of results worse
5	Stimulation	0,41	Bad	In the range of the 25% worst results
6	Novelty	1,00	Above average	25% of results better, 50 of results worse

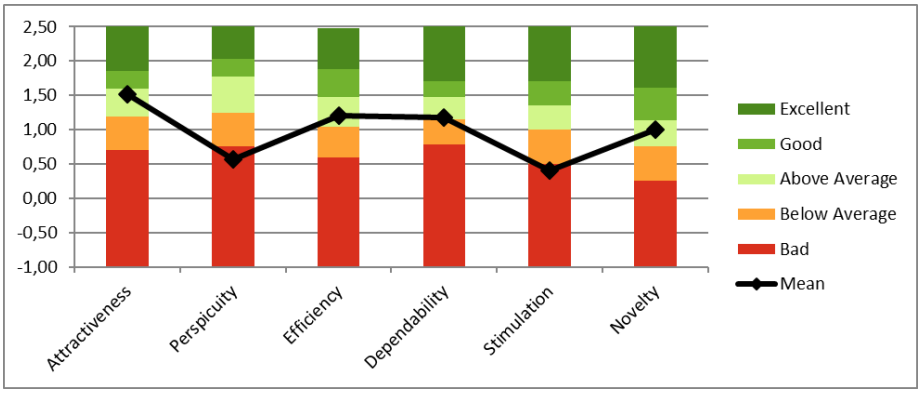

Figure 6. Graphic Bars of User Experience Questionnaire Score Results

Based on the results of the UEQ score in Figure 6, it can conclude that the programs built generally get an "Above Average" score. Thus the schedule is ready to be implemented for direct use in schools which are the object of this research.

3.15. Discussion Result

Based on the research that has been finishing, it has produced a program that can portray student's mental conditions, makes it easier for teachers to explore student problems, and provides predictions for students who have potential issues. From this study, the K-NN algorithm was implemented in a high school environment, wherein in the previous research, it was implemented in a college environment. It is hoping that the system built can improve the quality of education by grouping study groups based on the modality or tendency of student learning. This will make it easier for subject teachers to determine learning syntax. The output of student personality report cards can help students understand how they should learn and how to act in accordance with cyclical conditions.

4. CONCLUSION

After conducting research and evaluation, researchers can conclude that the application of the prediction system and student personality mapping is able to map the mental conditions of students where the data can be the basis for the preparation of counseling work programs, can assist teachers in making the right approach to students according to student personalities able to predict students with potential problems using the k-NN Algorithm with an accuracy rate of 83%. The results of measuring user experience based on a prediction system built using the User Experience Questionnaire (UEQ) concluded that the system reached Above Average.

5. REFERENCES

[1] C. J. L. Murray, A. D. Lopez, and World Health Organization, *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: Summary*. World Health Organization, 1996.

[2] I. Maulana, A. Sriati, T. Sutini, E. Widianti, I. Rafiah, and N. O. Hidayati, “Penyuluhan Kesehatan Jiwa untuk Meningkatkan Pengetahuan Masyarakat tentang Masalah Kesehatan Jiwa di Lingkungan Sekitarnya MKK : Volume 2 No 2 November 2019 Orang yang mengalami gangguan Jiwa di Dunia ini sudah banyak dan bahkan di Indonesia pun banyak p,” vol. 2, no. 2, pp. 218–225, 2019.

[3] A. Goodman, D. L. Lamping, and G. B. Ploubidis, "When to use broader internalizing and externalizing subscales instead of the hypothesized five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children.," *J. Abnorm. Child Psychol.*, vol. 38, no. 8, pp.

1179–1191, 2010, DOI: 10.1007/s10802-010-9434-x.
[4] D. Kurniadi, E. Abdurachman, H. L. H. S. Warnars, and W. Suparta, "The prediction of scholarship recipients in higher education using K-Nearest neighbor algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, no. 1, p. 012039, 2018, DOI: 10.1088/1757-899X/434/1/012039.
[5] D. Yonitasari and R. Setiyani, "Pengaruh Cara Belajar, Lingkungan Keluarga, dan Fasilitas Belajar Terhadap Prestasi Belajar Ekonomi Akuntansi Siswa Kelas XI IPS SMA Negeri 4 Magelang Tahun Ajaran 2013/2014," *Econ. Educ. Anal. J.*, vol. 3, no. 2, pp. 241–248, 2014.
[6] M. Mustakim and G. Oktaviani, "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa," *J. Sains dan Teknol. Ind.*, vol. 13, no. 2, pp. 195–202, 2016.
[7] H. Susanto and S. Sudiyatno, "Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu," *J. Pendidik. Vokasi*, vol. 4, no. 2, pp. 222–231, 2014, doi: 10.21831/jpv.v4i2.2547.
[8] R. Shalihah and Y. S. Nugroho, "Prediksi Perilaku Siswa SMK N 2 Surakarta Menggunakan Metode Naïve Bayes," Surakarta, 2016.
[9] A. Rohman, "Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa," *Neo Tek.*, vol. 1, no. 1, 2015, doi: 10.37760/neoteknika.v1i1.350.
[10] R. S. Pressman, *Software Engineering A Practitioner's Approach, Seventh Edition*. 2010.
[11] S. Aswati, M. S. Ramadhan, A. U. Firmansyah, and K. Anwar, "Studi Analisis Model Rapid Application Development Dalam Pengembangan Sistem Informasi," *J. Matrik*, 2017.
[12] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
[13] H. B. Santoso, M. Schrepp, R. Yugo Kartono Isal, A. Y. Utomo, and B. Priyogi, "Measuring user experience of the student-centered E-learning environment," *J. Educ. Online*, vol. 13, no. 1, 2016.
[14] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. 2014.
[15] F. Mangunsong, *Psikologi dan pendidikan anak berkebutuhan khusus jilid 1*. Jakarta: LPSP3UI, 2009.
[16] B. Walgito, "Bimbingan Konseling (Studi dan Karier)," *Yogyakarta: Penerbit ANDI*, 2010. .
[17] S. Sutirna, "Buku Bimbingan Konseling (Pendidikan Formal, Non Formal, dan Informal)," *Univ. Singaperbangsa Karawang*, no. March, 2019.