Counselor Application Frontend With Personality-Matching Using Android-Based K-Means Clustering Algorithm

Ifan Perdana Putra[1], **Eko Hari Rachmawanto***[2]**, Wellia Shinta Sari**[3]**, Tri Esti Rahayuningtyas**[4],
Choerul Umam[5]**, Mahadika Pradipta Himawan**[6]**, Noorayisahbe Bt Mohd Yaacob**[7]

[1,2,3,4,5,6]*Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah 50131*
[7]*Center for Software Technology and Management (SOFTAM), Faculty of Information Science and Technology, University Kebangsaan Malaysia (UKM).*
Selangor, Malaysia.
E-mail : 111202080012@mhs.dinus.ac.id[1]*, eko.hari@dsn.dinus.ac.id**[2]*,*
wellia.shinta@dsn.dinus.ac.id[3]*, tri.esti.rahayuningtyas@dsn.dinus.ac.id*[4]*,*
choerul.umam@dsn.dinus.ac.id[5]*, mahadika.pradipta@dsn.dinus.ac.id*[6]*,*
noorayisah@ukm.edu.my[7]
*Corresponding author

Abstract - Education is one of the most important things for people to have. Many people are competing for education to increase their abilities. Technology plays a big role in developing access to education to make it easier with many online applications and online classes education becomes easier. However, there are still many unresolved problems in this field of education, namely the emergence of the phenomenon of incompatibility between educators and students so that student interest decreases dramatically because of this. And also, the lack of learning materials taught that are not school subjects such as programming. Therefore, the author and team designed an application where this application can find students a learning mentor outside of school so that they can increase their knowledge. The application also provides a matching feature based on the student's personality so that the student can find a suitable tutor.

Keywords - personality, application, android, front-end.

1. INTRODUCTION

Education is a basic thing that must be owned by someone, because education makes someone have their own skills and creativity that stand out and can be used in life. Because of the importance of education, all people must get this education, both formal and informal education. Education allows a person to be able to develop his potential from personality, self-control, culture, morals and the nation state [1]. All of the things mentioned above can be obtained by someone with adequate and sufficient education. The things above have an important role in today's life and can only be obtained through education.

Technology is now developing faster and bigger. education is also one of the factors that is touched by this very rapid technological development. technology supports the educational process as a tool and enhances learning. the existence of gadgets such as smartphones, computers, laptops and other technologies makes education now easily accessible [2]. The role of the internet in finding information and knowledge is also very large in education. The existence of virtual conferences such as Google Meet and Zoom makes it easy for teachers and students to communicate remotely feels easier in the teaching and learning process. Online

tutoring applications such as Ruangguru and Zenius are also very helpful for students in understanding school lessons better. The use of technology in education also developed more rapidly during the Covid-19 pandemic, where all people were prohibited from leaving the house to maintain their health. Covid-19 forced people to use technology as a means to guide knowledge and get an education [3].

Although many educational problems are solved by the above technological developments, some problems still arise related to education itself [3]. One of the problems is the wrong field of study, the lack of out-of-school education and also the incompatibility of educators with students so that students' interest in the material provided is reduced.

As said above, the problem of media to develop students' interests and talents outside of school is still very minimal, especially in Indonesia. The development of interests and talents is very important and influential in students [4]. By developing the interests and talents that students want, students will feel happier and will also feel an increase in self-confidence. As well as an increase in the percentage to be accepted in the appropriate job field. Unfortunately, there are still not many media to express themselves. The applications mentioned above such as Ruangguru and Zenius are still in the form of school materials and there are still no materials for developing interests and talents.

The next problem that is still widespread in the world of education is the problem of incompatibility between teachers and teachers in the scope of teaching and learning [5]. In research in human relations, individuals will be more compatible and have more inner bonds if they have relationships with individuals who have the same personality. In education, students will often meet with teachers who do not have the same personality as him so that he will tend to get bored and ignore the lessons delivered by the teacher so that the knowledge gained is not conveyed.

Seeing the above problems in education, the author's team has a plan to provide media where students will be given the convenience to not only learn according to the interests and talents, they have by being given the freedom to choose the desired education subject, but students can also have the freedom to find the right teacher and according to their respective personalities [6]. This media is an android-based application that allows students to hire a personal tutor who can teach what students want and can use the personality matching feature.

This personality matching will use K-means clustering to group the personalities of teachers and students and Euclidean distance to find the teacher with the highest percentage. K-means clustering allows clustering with a certain constant so that it is easier to implement with limited clustering such as personality and also clustering many clusters [7].

In the project, the author is tasked with creating all parts of the android front-end, which includes all parts of the interface with the user and also integration with the back-end that provides and stores data in this application. the author will use Android Studio to create a front-end display with the Jetpack Compose library [8]. As well as the Agile Model as a way of developing software. This model is used so that the course of development is organized by iteration, so that if there are features that can be added it will be easier to implement.

2. RESEARCH METHOD

2.1. Data Collection Method
Data collection in this study used three events, namely:
1. Literature Study

    In the literature study, some relevant application development literature will be collected where the data to be collected is data on how other researchers make the same application, what development frameworks are used, user experience data for

making a satisfying user interface and advanced data on how the application will be developed in the future. This literature will be searched in journal collection websites such as google scholar, sci-hub and so on.

2. Interview

Interviews are a way of collecting data where researchers will ask questions to sources related to the research being conducted. In this interview the author conducted interviews with several prospective application users. The data desired from this interview is data on the features that users need, how important features are in the application and input features for the future.

3. Observation

In observation, researchers will observe in the form of several schools and / or campuses to find data on what study programs and lessons are in schools and campuses, as well as observing other applications where in the application will be observed what out-of-school lessons can be used in the application. The desired data is data on lessons and study programs and what out-of-school lessons are the most popular among teenagers now.

2.2. Data Analysis

After the data is collected all data will be analyzed and conclusions will be drawn from the analysis. To test the validity of existing data in the literature study, the author will compare data from several articles on the same topic, some of the information will be validated and checked for validity with other literature [9]. In some literature the application has a lot in common, namely on the android platform, this is because android is now the most widely used operating system by all people around the world, therefore the application has a large marketing if developed in this android platform, therefore the author will use android as an application development platform [10]. Applications from literature also use a lot of frameworks that are used, the author chooses the agile framework, especially the kanban development framework is very flexible for application development and also the application will be easier to develop and raise in the future.

In the results of interviews conducted, prospective users want to get lessons that vary unlike schools or campuses that have only such lessons. Prospective users also support the application of mentor matching so that the teaching and learning process is more enjoyable and learning materials can be given more attention.

In the observation results it turns out that the lessons in schools and campuses have very many charts in the faculty of computer science also have very many lessons. In the application of learning-related applications used are also very much and varied such as sports, music, technology and others [11]. So the author will make a flexible application where the mentor will create a table of their abilities in teaching.

2.3. Proposed Method

In this application development method will use agile methods for application development frameworks, where the agile method used is kanban development. Where this development emphasizes the development of repeated cycles so that the application has the desired features and also the application can be developed and enlarged more easily in the future. unlike traditional development this method emphasizes flexibility and prioritizes time.

In this phase the developer will determine the need to explain the business opportunities and time and effort to develop the application. Where in this case we will create an android application where this application will provide a place to rent a learning mentor with personality matching [12].

1. Requirement Gathering and Analysis

   In this phase the developer will determine the need to explain the business opportunities and time and effort to develop the application. Where in this case we will create an android application where this application will provide a place to rent a learning mentor with personality matching.

2. Design Requirement

   The author will develop the application design according to the planned features in the application such as search pages, home pages, teacher pages, teacher detail pages, schedule history pages, rental history pages, profile pages, matching survey pages and others using user interface design and user experience.

3. Construction/Iteration

   In this stage the developer starts making applications that are made with improvisation of minimal features. And will be iterated according to testing feedback.

4. Deployment

   The developer will implement the application as a whole

5. Testing

   Developers will conduct unit tests and black box tests and also look for bugs in the application. With in-app tools and testers.

6. Feedback and Maintenance

   After testing and the application will be carried out maintenance of this android application and also development according to the previously existing testing input.

2.4. Machine Learning Model Data Collection

Data collection for machine learning models is done in two ways, namely:

1. Sample Data

   The author uses data samples from the Kaggle website in the form of data results from big five personality questions containing answers to 50 questions about personality.

2. Survey

   The author also conducts a separate survey with the same questions from the data sample to be used as testing data.

2.5. Machine Learning Model Data Analysis

The data that has been obtained will be processed to make the data into the desired model. this processing will be carried out in three stages, namely:

1. Clustering stage

   In this clustering stage, sample data from the Kaggle website will be used to cluster the answers into several clusters to be simplified for matching with the k-means clustering method. After clustering all data will be labeled with a personality.

2. Classification Stage

   Data that has been labeled will be classified where this stage will be used to classify new survey data for users who have just conducted a survey. Labeled data will be split from data and labels from clustering to be split again as training data and testing data.

3. Matching Stage

   In this matching stage, survey data will be used and classification will be carried out with the classification model that has been made as supervisory data. For matching the data to be used are gender data, categories of mentors and user personality survey data.

2.6. Methods in Machine Learning Model

The methods that will be used in machine learning are as follows:

1. Clustering Method

    In the clustering stage, the method used is k-means clustering. From the Kaggle data sample, clustering will be carried out with five clusters for classification simplification. And the data that has been clustered will be labeled according to the data.

2. Classification Method

    In the personality classification stage, the labeled data will be classified using TensofFlow DNN where the data will be separated from the label and then divided into training data and testing data to create a model that will be used to classify new user data.

3. Matching Method

    In matching the data from the survey will be entered into the personality classification and stored as mentor data. After that, new user data will also be entered into the personality classification after that the classification data will be compared with the mentor data and the closest data will be found from the user data with Euclidean distance.

2.7. Problem Solving Evaluation

    In the research evaluation of the application will solve the above problems can be measured by the following user satisfaction parameters that will be carried out to evaluate the problem solving:

1. Conduct surveys and interviews with a number of students and students related to the problems above.
2. Screening users from surveys for the suitability of users needed.
3. Testing the application to these users and recording the results.
4. Conduct a second survey where the testing users indicate their satisfaction in using the application and whether the application solves their problems.

3. RESULTS AND DISCUSSION

3.1. Machine Learning Model Results

    The model uses 3 stages: clustering, classification, and matching. In clustering the data will be clustered and labeled to be used as classification material. The following are the results of K-Elbow Visualization to find clusters and also the results of clustering data:
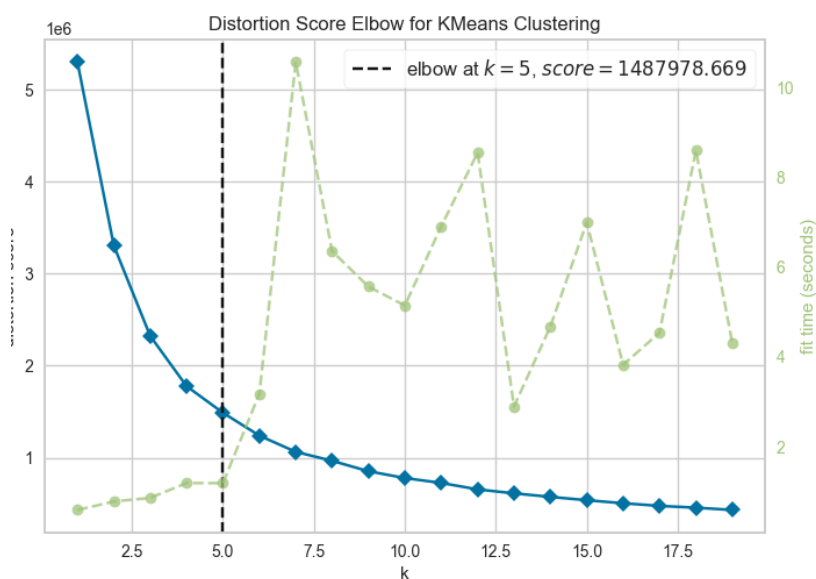
Figure 1. K-Elbow Visualisation

In the visualization above, the data will use 5 clusters for data clustering. After knowing the optimal cluster, the data will be clusterized using K-Means with 5 clusters. The following are the results of clustering the data:
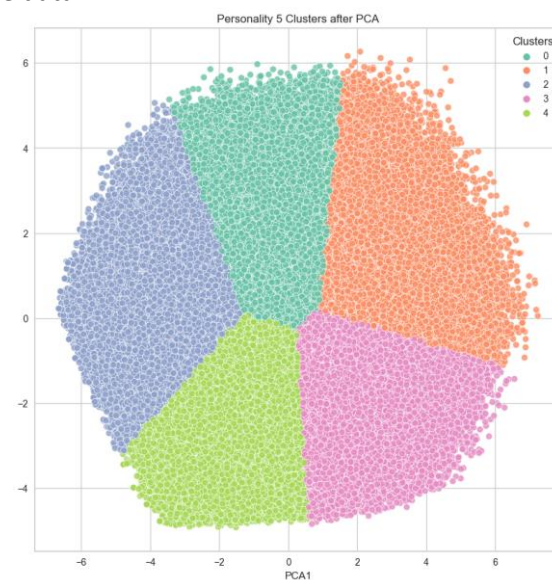

Figure 2. Data Clasterisation Visualized

The visualization above shows the clustering results of all existing data. Finally, the data will be labeled according to the cluster and will be used for the personality classification model.

In the classification stage, a model will be created where the model can label the user's answers to questions about personality with labels that have been created from the clustering stage with data that has been labeled. Labeled data that comes from clustering will be used for training the model so that the accuracy of the model is high to classify new data. Here are the results of the training data:
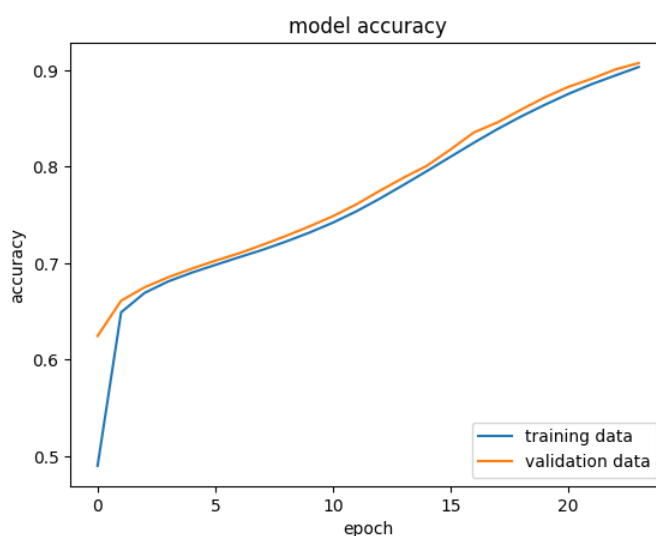
Figure 3. Accuracy Increase During Training

The visualization above is a visualization of model accuracy when training is carried out. It can be seen that the model shows an increase in accuracy. Accuracy reaches 91%. After the training model is complete the model will be tested to see the precision of the model in the testing. The following are the results of testing data:

```
predicted_clusters[:20]
```

```
array([3, 0, 2, 3, 4, 4, 1, 3, 0, 4, 4, 0, 1, 0, 1, 4, 0, 0, 0, 3],
      dtype=int64)
```

```
# Calculate precision
precision = precision_score(y_test, predicted_clusters, average='micro')

# Calculate recall
recall = recall_score(y_test, predicted_clusters, average='micro')

# Print the results
print("Precision:", precision)
print("Recall:", recall)
```

```
Precision: 0.9074111789203121
Recall: 0.9074111789203121
```

Figure 4. Results of Model Testing

It can be seen that the model can achieve a precision of up to 0.90 which is considered high. Furthermore, for the matching stage, the classification model will classify the personality of survey answers made for dummy tutor data and also survey data from new users. After that, other data such as gender and category are added, these data will be used to match new users with existing tutor data by simplifying the data into 2 important data with PCA and then looking for the proximity of user data to tutor data and tutor data will be searched for the top few closest to user data using Euclideans Distance. Here are the results of the data matching:

```
User Index: 0                    User Index: 58
Matched Peers:                   Matched Peers:
Index: 81 | Score: 80.94%        Index: 5 | Score: 63.40%
Index: 3 | Score: 79.42%         Index: 25 | Score: 63.40%
Index: 95 | Score: 47.57%        Index: 8 | Score: 20.61%
Index: 28 | Score: 36.41%        Index: 92 | Score: 20.61%
Index: 94 | Score: 36.41%        Index: 13 | Score: 18.30%
Index: 34 | Score: 31.73%        Index: 57 | Score: 12.36%
Index: 41 | Score: 12.38%        Index: 83 | Score: 0.00%
```

Figure 5. Matching Results

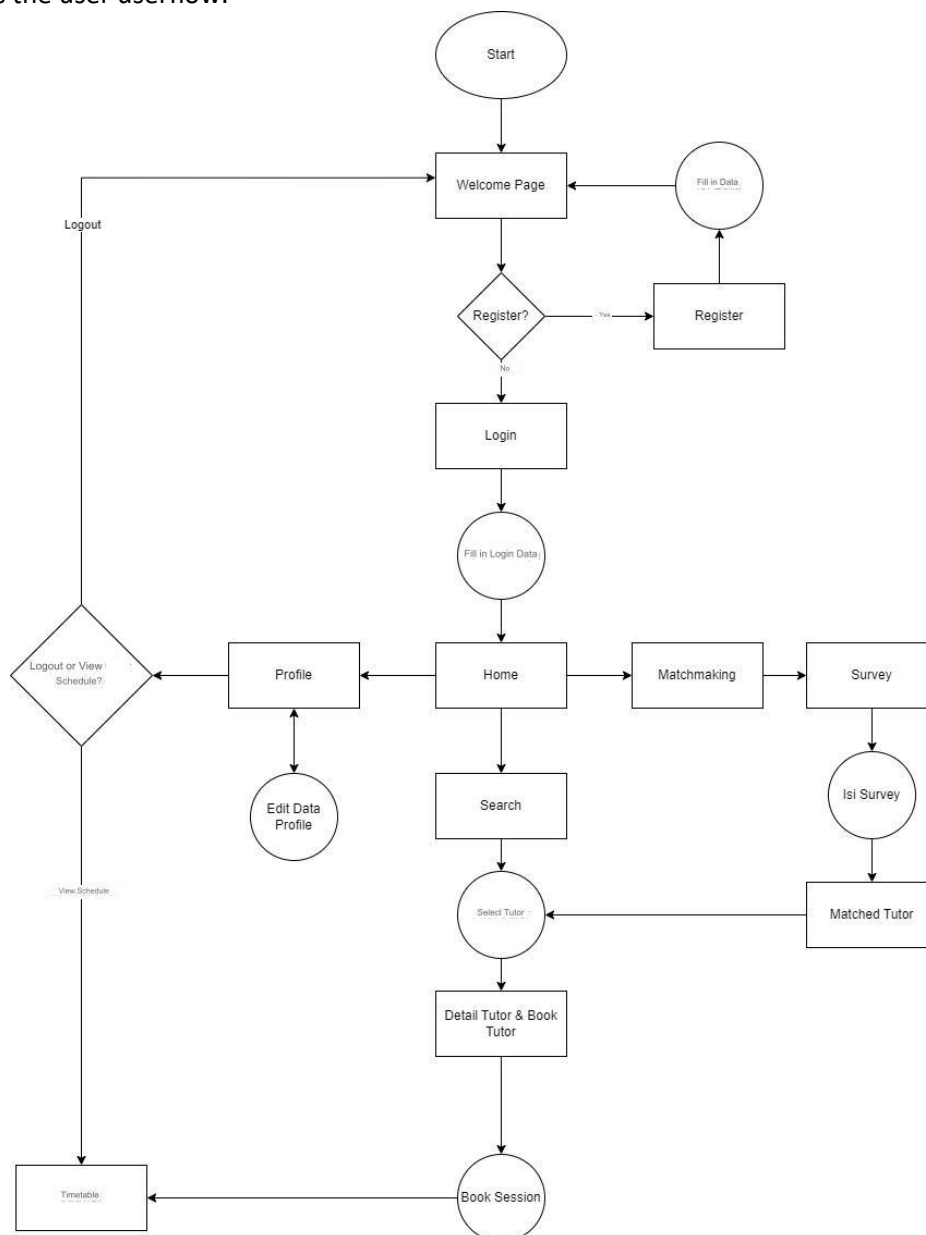## 3.2. Results of Application Development
Below is the user userflow:

Figure 6. Flowchart

36

The userflow above is used to understand the user's path from start to finish.
Furthermore, making applications with agile methods will be separated into two iterations. The results of the first iteration are as follows:
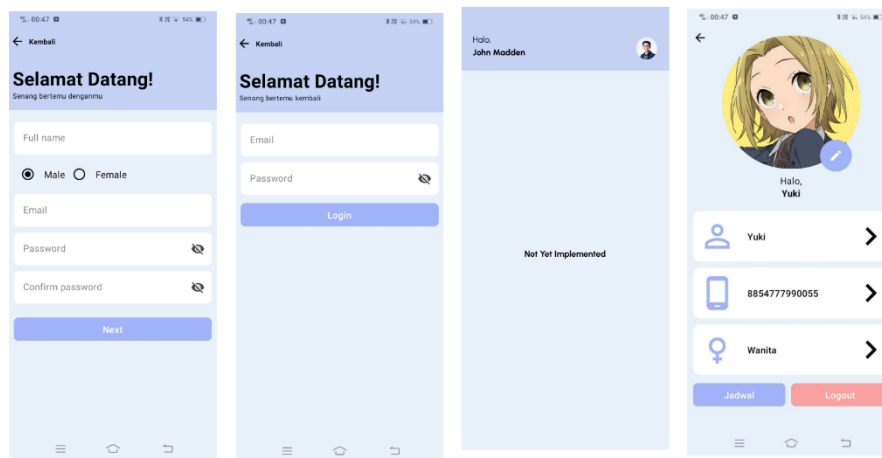

Figure 7. First Iteration Results

In the first iteration, user interactions such as login, register and user profile change were created. Furthermore, in the second iteration, the following are the results:
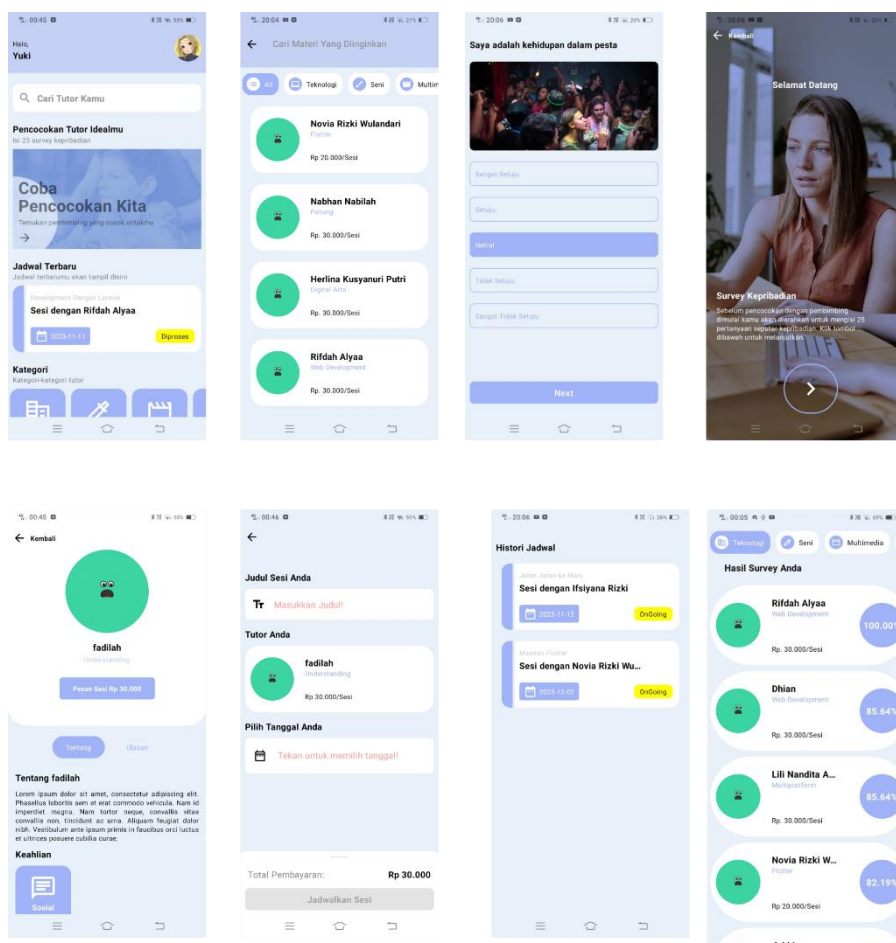

Figure 8. Second Iteration Results

The second iteration created all the things that users can do such as adding mentors and so on. Both have passed unit testing and user testing.

3.3. User's Feedback

The survey of the application has a result that 100% of users can navigate well, edit and add well, 90% of users are positive that the application will offer many tutors with different specialties with the specialization and category input so that users can hone any material they want with this application, and 75% of users believe the use of matching will make teaching and learning activities more exciting with suitable tutors and the user's own choice. Feedback given by users is mostly about UI improvisation and also additional features that will be done in further development.

4. CONCLUSION

4.1. Conclusion

Based on the results of the research that has been done, several conclusions can be drawn, namely:
a. Research produces freelance applications for supervisors out there to do teaching, as well as for users who are looking for their own mentors according to their wishes and can use matching using personality surveys to match their personality with existing mentors.
a. Based on testing by means of surveys to several users, the application can solve problems where the application has a variety of teachers with varied material and teachers can input themselves so that users can hone their interests. And can alleviate the problem of incompatibility of teachers and students by the way students can find their own mentors and can also use personality matching.

4.2. Suggestion

In further research, it is hoped that research can be continued by developing the above application, because the application is still in the MVP stage and there is a lack of applications for mentors, it is hoped that in the future other features can be added and also applications for mentors can be made. As well as from the feedback survey from users can be resolved such as the lack of UI appearance and also additional features.

REFERENCES

[1] M. Mardi and M. R. Keyvanpour, "GBKM: A New Genetic Based K-Means Clustering Algorithm," in *2021 7th International Conference on Web Research (ICWR)*, Tehran, Iran, 2021.

[2] V. Venkatesh and A. S. Vindhya, "Predicting the Accuracy of Fractionation of Patron's Activities in Online Social Networks Using Novel K-Means Clustering Algorithm Comparing with Agglomerative Hierarchical Clustering Algorithm," in *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, Gautam Buddha Nagar, India, 2023.

[3] S. Bhatia, "New improved technique for initial cluster centers of K means clustering using Genetic Algorithm," in *International Conference for Convergence for Technology-2014*, Pune, India, 2015.

[4] W. Yintong, L. Wanlong and G. Rujia, "An improved k-means clustering algorithm," in *World Automation Congress 2012*, Puerto Vallarta, Mexico, 2012.

[5] J. Li, W. Liu, M. Liu and M. Huang, "Study on Chinese Text Clustering Algorithm Based on K-mean and Evaluation Method on Effect of Clustering for Software-intensive System," in *2020 International Conference on Computer Engineering and Application (ICCEA)*, Guangzhou, China, 2020.

[6] A. Alam and M. Muqeem, "Automatic Clustering for Selection of Optimal Number of Clusters by K-Means Integrated with Enhanced Firefly Algorithms," in *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, 2022.

[7] A. Sen, M. Pandey and K. Chakravarty, "Random Centroid Selection for K-means Clustering: A Proposed Algorithm for Improving Clustering Results," in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Gunupur, India, 2020.

[8] M. Run-Yang, X. Feng-Guo, Y. Feng-Huang and B. Quan-Chen, "BPNN anomaly data detection algorithm based on gray wolf algorithm to optimize K-means clustering," in *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, Hangzhou, China, 2021.

[9] N. D. Sugiharto, D. Elbert, J. Arnold, I. S. Edbert and D. Suhartono, "Mall Customer Clustering Using Gaussian Mixture Model, K-Means, and BIRCH Algorithm," in *2023 6th International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, 2023.

[10] J. Dan, "Research and Improvement of K-means Clustering Analysis Algorithm in the Information Warfare," in *2022 3rd International Conference on Computer Science and Management Technology (ICCSMT)*, Shanghai, China, 2022.

[11] R. Sun, "Design Optimization of Integrated Energy System Using Virtual Energy Storage under the Use of K-means Clustering Algorithm," in *2022 2nd International Signal Processing, Communications and Engineering Management Conference (ISPCEM)*, Montreal, ON, Canada, 2023.

[12] L. Bai, Z. Song, H. Bao and J. Jiang, "K-means Clustering Based on Improved Quantum Particle Swarm Optimization Algorithm," in *2021 13th International Conference on Advanced Computational Intelligence (ICACI)*, Wanzhou, China, 2021.