International Journal of Engineering, Science and Information Technology
Volume 5 No. 3 (2025) pp. 41-50
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i3.879
Research Paper, Short Communication, Review, Technical Paper

Clustering of Futsal Interest Level Among Students K-Means Method

Faris Bagaswara*, Muchlis Abd Muthalib, Rini Meiyanti

Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

**Corresponding author Email: farisbagass9@gmail.com*

The manuscript was received on 25 November 2024, revised on 2 January 2025, and accepted on 1 April 2025, date of publication 11 May 2025

Abstract

Futsal is a small field sport with a time of 20 minutes per round. Malikussaleh University is one of the universities that initiated Futsal as a health sport for its students. To determine students' interest in Futsal, clustering was carried out using the K-Means method on 100 students of the Faculty of Engineering involved in this study. This research proposal uses five variables: time variables, field facilities, motivation, environment, and plans. This study aims to help students at Malikussaleh University of Engineering find out what level of interest students have in Futsal. Grouping is based on data mining to determine the pattern of each sequence. Data mining includes tracking patterns, classification, association, outlier detection, clustering, regression, and forecasting. This study also led to an innovative grouping system using the Python programming language and MySQL as a database. The K-Means Clustering algorithm used in this grouping system states that out of 100 Malikussaleh University students, 20 people are students who have a professional player futsal interest level (C1), 28 students have a regular player futsal interest level (C2), five students have a Beginner player futsal interest level (C3), 47 students have an amateur player futsal interest level (C4). The study results showed that 20% were professional, 28% were regular, 5% were beginner, and 47% were amateur players. These results indicate that the interest in Futsal for Malikussaleh University students is still minimal, so encouragement is needed for students to participate in futsal activities.

Keywords: *Clustering, Futsal interest, Students, K-Means Method.*

1. Introduction

Futsal is one of the most popular and in-demand sports. This is because Futsal is easy to play and does not require a large field. Futsal refers to Spanish, separated into "futbol" and "sala". Futsal is a soccer game, with a smaller field and goal; each team consists of five people. In the game of Futsal, the ball must be put into the opponent's goal as many times as possible to win. A player can score a goal by kicking the ball into the goal, heading it, or using other body parts that are allowed to touch the ball. This game is played with fewer people than soccer. The time used is also not too long, consisting of two rounds lasting 20 minutes. As an indoor soccer sport, Futsal has become an increasingly popular activity among students. This activity provides a means of recreation and promotes a healthy lifestyle and social interaction among students. At Malikussaleh University, interest in Futsal is considered one of the indicators of student health and fitness. Therefore, it is necessary to analyze five (5) variables: time, field facilities, motivation, environment, and plans. The time variable sourced from the website has four categories: low, low-medium, medium, and high. The field facilities variable sourced from Google has four categories: less good, quite good, good, and excellent. The motivation variable has four categories: entertainment, exercise, socialization, and competition. The environment variable has three categories: inactive, quite active, active, and very active. The plan variable sourced from the website has two categories: there and there.

This study focuses on male and female students studying at Malikussaleh University. Many male and female students at Malikussaleh University do not know their interest in Futsal. To help male and female students at Malikussaleh University know at what level their interest in Futsal is, a system was built using one of the data mining classifications, clustering, according to how it works. Data Mining is an iterative process aimed at database analysis that seeks to filter information and knowledge that can prove data accuracy, potentially for experts involved in decision-making and problem-solving. In the simplest definition, Data Mining automatically recognizes related patterns in a database [1]. Clustering is the process of grouping a set of data objects (into multiple groups) or clusters so that objects in a group have high similarities, but are very different from objects in other groups [2]. The method applied in this study uses clustering

of futsal activity interests, which will be carried out, namely, the k-means method. K-Means is a data grouping technique where the existence of each data point in a cluster is determined by the degree of membership. This technique was first introduced by Jim Bezdek [3], [4].

The k-means algorithm begins with a random selection of K, the number of clusters to be formed. Then the K value is randomly determined to be the cluster's center (Centroid) so that the closest distance from each data point to the Centroid is found. The location of this research will be the Malikussaleh University campus environment.

A. Muh conducted previous research. Ismail [5] with the percentage of interest data results on 40 students with an outstanding category of 20.0%, a good category of 37.5%, a moderate category of 17.5%, a less than ideal category of 12.5%, and a significantly less than perfect category of 12.5%. Atang Supriatna et al [6] Students 1105 data results obtained 26 Clusters. Mardiansa et al. [7] Science subject student score data obtained from clustering results showed that high student interest was 41.9%, moderate student interest was 54.8%, and low student interest was 3.2%.

Futsal was popularized in Montevideo, Uruguay, in 1930 by Juan Carlos Ceriani [8], a soccer coach from Argentina. He got the idea when his soccer team wanted to practice on the soccer field, but they couldn't do it because of the rain. Ceriani took the initiative to move his soccer practice indoors [9]. FIFA issues futsal rules, which include several significant differences from conventional soccer. The futsal field is smaller, the game time is shorter, and there are fewer players. These rules are designed to increase the speed and intensity of the game.

2. Research Methods

2.1. Data Mining

As a relatively new field of science, data mining is currently one of the centers of attention among academics and practitioners. Data mining is the process of obtaining useful information from large databases that need to be extracted to become new information and can help in decision-making [10]. Data mining is the discovery of significant patterns from large data sets, which can be stored in various storage locations such as databases, data warehouses, or other information storage systems. This activity closely relates to multiple disciplines, including database systems, data warehousing, statistics, machine learning, information retrieval, and advanced computing. Data mining is used to build models and patterns that can be used to make predictions about data based on previous data in a specific period. In general, data mining functions as a tool to extract knowledge from vast volumes of data. The basic concept of data mining is determining hidden information in a database. It is part of Knowledge Discovery in Databases (KDD) to find helpful information and patterns in data. Data mining searches for new, valuable, and helpful information in a data set involving computers and humans. It is iterative, either through automatic or manual processes [11]. Data mining is the process of extracting meaningful information from large data sets. In computer science, data mining is the process of finding hidden patterns to generate new knowledge from existing data. Activities related to data mining can be grouped into four main categories: prediction modeling, cluster analysis, association analysis, and anomaly detection.

Predictive modeling is concerned with creating a model that can map each set of variables to each target, and then the model is used to assign target values to the new set obtained.

2.2. Tracking Patterns/Sequencing

The first data mining technique is tracking patterns or sequences of events. This technique finds a pattern in a series of sequential events. The tracking patterns technique can detect something at specific intervals, such as a spike in product demand during the weekend or the number of people visiting the site during certain weather conditions [12].

2.3. Classification

Classification techniques require more complex data mining techniques because they require collecting all data from a particular class or category. This technique is the most commonly used. Because it can be applied to group data based on the desired label. For example, based on financial information and transactions, customers can be categorized into low, medium, or high credit risks [13].

2.4. Association

The following data mining classification is an association, market basket analysis, related to product marketing. Basket analysis aims to identify products that customers often purchase together. For example, a customer may buy a packaged soda drink when buying a particular snack. Knowing customer habits like this, the company can label certain products as "people also bought this" on the marketplace [14].

2.5. Outlier Detection

This technique aims to identify when there is an anomaly in the data pattern. For example, a product is usually always purchased by male customers, but one week in February, there is a sudden spike in purchases made by female customers. The outlier detection technique plays a role in analyzing the spike and its causes to decide the next sales steps [15].

2.6. Clustering

Clustering techniques are similar to classification but require more labels or groups of data based on similarity patterns. For example, you want to group different audience demographics into groups based on their background, finances, or the number of purchases when shopping at a particular store. Clustering analysis is the process of dividing data in a set into several groups whose data similarities within a group are greater than the similarity of the data to the data in other groups. The potential of clustering can be used to find the data structure that can be used further in various applications such as classification, image processing, and pattern recognition [16].

Clustering is a technique in data mining that aims to extract interesting patterns from large data sets. In simple terms, clustering can be interpreted as grouping data into several groups or clusters, where the data in one cluster has high similarities. Several approaches are used in developing clustering methods, with two main approaches: partition-based and hierarchy-based. Clustering with a partition

approach, often called partition-based clustering, is grouping data by sorting the analyzed data into several existing clusters. This does not take into account the hierarchy of the data. In the partitional clustering method, each cluster has a cluster center point (Centroid), and in general, this method has a goal function of minimizing the distance (dissimilarity) from all data to the center of each cluster [17].
Clustering with a hierarchical approach, often called hierarchical clustering, groups data by creating a hierarchy in the form of a dendrogram (a type of tree-like diagram used in hierarchical grouping) where similar data will be placed in adjacent hierarchies and those that are not in distant hierarchies. In hierarchical clustering, data is grouped through a chart as a hierarchy, where the two closest groups are merged in each iteration, or the entire data set is divided into clusters. The steps to perform hierarchical clustering are to identify items with proximity, then combine the items into one cluster, calculate the distance between clusters, and repeat from the beginning until all are connected.
Optimal clustering results will show a high level of similarity within a group, while the similarity between groups should be low. This similarity is measured numerically between two objects, where the similarity value is higher if the objects being compared have remarkable similarity. Conversely, the similarity value will be low if the objects differ. The quality of clustering is greatly influenced by the method used. In the clustering process, four data types are commonly used: interval scale variables, binary variables, nominal variables, ordinal variables, and ratio variables, as well as other data types.

2.7. Regression

The following data mining classification, regression, aims to find patterns in numerical values rather than classes. The result of this technique is a function that serves as a determinant based on the input value [18]. For example, it determines a product's price based on other factors such as availability, customer demand, and competitors.

2.8. Forecasting

The last data mining technique is the most valuable because it aims to predict the value achieved in a specific period. The prediction technique uses noisy data and values from the previous period as a reference or basis for prediction. For example, based on last month's purchase data, customers' purchases can be predicted for next month.
Since data mining is a series of processes, it is divided into several steps. These phases are interactive; users are involved directly or through intermediaries [19].

2.9. K-Means

K-Means is a data analysis or mining method that performs the modeling process without supervision (unsupervised). It is one of the methods that performs data grouping with a partition system. The k-means method attempts to group existing data into several groups, where data in one group have the same characteristics as each other and have different characteristics from the data in other groups. In other words, this method attempts to minimize variation between data in a cluster and maximize variation with data in other clusters.
K-Means is one of the non-hierarchical data clustering methods that attempts to partition existing data into one or more clusters, so that data with the same characteristics are grouped in the same cluster. Data with different traits are grouped into another group. The method is included in the distance-based clustering algorithm that divides data into several clusters, and this algorithm only works on numeric attributes [20].
K-Means is grouped using the cluster analysis method, which is directed at partitioning. N observation objects into K groups or clusters, where each observation object has a group with an average or mean. The k-means algorithm is included in the application of data mining clustering, which can be used to group data into several groups. The groups formed have predetermined criteria, and then the data that matches the group is collected into one cluster. The center point, or Centroid, is something that each cluster has [21].
The stages of the k-means algorithm are as follows:

1. Determine the number of clusters K to be formed (the value is free).
2. Select random data from the existing data, such as many K values, to be used as the initial Centroid.
3. Calculate each object's distance to each cluster's Centroid based on the Manhattan distance.
4. Allocate each data point based on its proximity to the Centroid (smallest distance).
5. Determining the new Centroid
6. Return to step 3 if the position of the new Centroid is not the same as the old Centroid.

2.10. Normalization

Normalization is one of the initial processes of a calculation procedure in data mining, where data attributes are scaled to obtain a specific small range, such as from -1.0 to 1.0 or 0.0 to 1.0. [22]. In their study on extensive database management, the team argues that normalization remains an essential step in data management, even though new technologies such as NoSQL and distributed databases are becoming more popular in certain cases. They suggest that denormalization may be a better alternative for some instances when database performance is more critical than reducing data redundancy [23].

2.11. Min-Max Normalization

Min-max normalization is a normalization method that performs a linear transformation on the original data to balance the comparative values between the data before and after the process [24]. Min-max Normalization is one of the data standardization methods carried out so that the data is in a smaller interval, such as [-1,1] or [0,1], so that no data is too large or too small to facilitate analysis. This method changes data values into values on the desired minimum and maximum scale.

2.12. Website

A website is a medium consisting of several interconnected pages, and it functions as a medium to display information, either in the form of images, videos, text, sound, or a combination of all of them. The website is multi-platform, which means it can be opened from any device connected to the internet. Although this technology has been used for a long time, many companies still use websites to display company profiles, sell products, or as a system that customers can use [25]. A website is a collection of pages in one domain, presenting various information that internet users can access and view through search engines. The information presented usually

includes images, illustrations, videos, and text for multiple purposes. Usually, the main page (homepage) can be accessed through a browser by entering the correct URL. This main page also contains several related pages, including the pre-production stage, which carries out the initial process for creating a website. At this stage, data collection is carried out either from clients or from literature with data in the form of text, images, videos, sound, and so on. At this stage, a planned concept is also carried out, such as determining the target audience, the website's theme, team planning, features that will be applied, and budget planning for the design schedule. For the production stage after entering the production stage, the data organization process is carried out here. A comprehensive visual web model is also created via a computer at this stage. Visual modeling in web design can be integrated into diagrams and so on. At this production stage, all owned data will be incorporated into a web format through programming using the Hypertext Processor (PHP) programming language and MySQL as its database management. The production stage has passed the pre-production and production phases, and then the post-production stage. At this stage, a trial of the web system that has been built is carried out by uploading it to web hosting and being called through a specific domain name, such as Python or MySQL.

2.1.3. Python

Python is a programming language that can execute instructions directly (interpretive) using an object-oriented approach. Known as an easy-to-learn language, Python was developed by a Dutch programmer, Guido Van Rossum. Many consider Python to be a programming language that offers a variety of capabilities, such as feature integration, clear and easy-to-understand syntax, and an extensive and comprehensive standard library. Although Python is a high-level programming language, it is designed to be easy to learn and understand for beginners. Python has commands that allow for easy and fast theory testing. Users can also choose, use, and enjoy simple libraries with a straightforward interface. This allows users to understand simple statements such as cycles and conditions. However, the main results of course depend on a person's desire to learn programming and computer science [26].

2.1.4. MySQL

MySQL is a database software. MySQL is a relational database that stores its data in interconnected tables. The advantage of storing data in a database is the ease of storing and displaying data because it is a table.[19]. SQL is a language used to access and retrieve data from relational or structured databases in the programming world. Meanwhile, MySQL is a database management system (DBMS) that uses SQL as a connecting language between software applications and database servers. MySQL is an open-source DBMS with two types of licenses: Free Software (free software) and Shareware (proprietary software with usage restrictions). With the GNU General Public License (GPL), MySQL can be used for free, personal, and commercial purposes, without a license [27].

2.15. Flow chart

In the development of each system, a series of stages must be passed, and one of them is creating a logical design for the system to be built. This logical design is usually realized in the form of a system design. One common form of system design is a flowchart or flow diagram, which graphically describes a program's steps and sequence of procedures. Flowcharts help in further understanding and evaluation of specific problem-solving [28].

2.16. Unified Modeling Language (UML)

UML (Unified Modeling Language) is an industry standard language to describe needs, conduct analysis & design, and determine architecture in object-oriented programming [29]. Unified Modeling Language (UML) is a standard object-oriented visual modeling language that provides benefits in the industrial world as a tool for visualizing, designing, and documenting software [30].

2.17. Activity Diagram (Activity Diagram)

Activity diagrams are one of the many types of visual modeling in the form of UML diagrams. An activity diagram is a diagram used as a tool to model or explain a process flow carried out in a system. Like a process sequence, this activity diagram is visualized based on the sequence of processes that occur in the system [31].

2.18. Research Workflow

Several steps must be taken regularly to carry out the research according to its objectives. The research steps are shown in Figure 1.

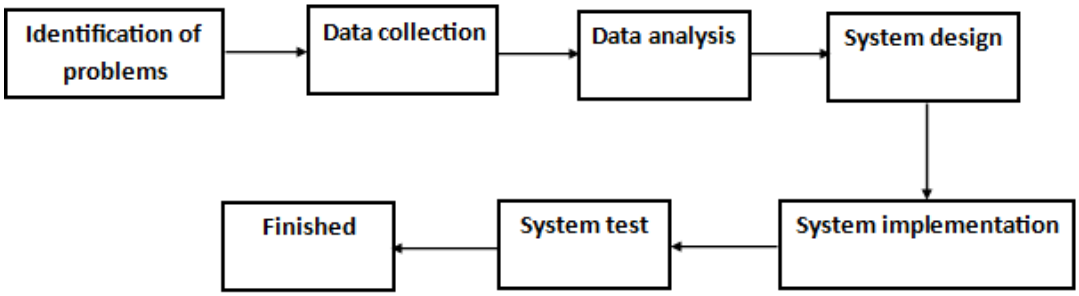

Fig 1. Research flow diagram

Based on the image above, we can see the steps of this research, which will be described as follows:
Identification stage: The problem is the identification of clustering using K-Means for futsal games. In the collection stage, data analysis is carried out to obtain and collect data/information needed to create a system to achieve research objectives. Analysis stage: Data collection uses several techniques or methods for securing information, library data, or relevant references. Design Stage The system plays a vital role because this is where the representation is carried out for the system that is implemented into a system workflow or

system scheme, so that with the system design, it will be more focused on developing and implementing the system later. At this stage, the author carries out the design process from raw data that will be processed into quality data. Data was obtained from an online questionnaire distributed to Malikussaleh University students interested in Futsal. Implementation stage: The designed system is applied to students to see the interest of futsal players in the research object. The system test stage tests students at Malikussaleh University to determine their interest in futsal sports. Testing phase: This will be done after the system is developed. This test aims to ensure whether the system that is built is running well and meets the needs or expectations in carrying out its functions, or if there are still inconsistencies with the system's design so that improvements need to be made again.
The Clustering System Scheme for the Level of Interest in Futsal in Students Using the K-Means Method can be seen in Figure 2.

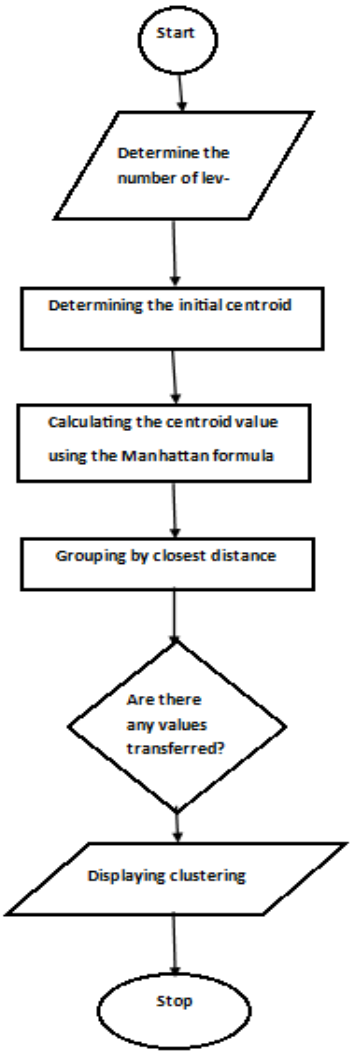

Fig. 2 System schematic diagram

In the initial process of the k-means method flowchart, the first step that must be taken is to input the questionnaire data on the level of interest in Futsal in students, determine the number of clusters, weighting rank, maximum iteration, and the most minor expected error, determine the initial Centroid based on the number of clusters to be formed, calculate the centroid value with the euclidean formula, group based on the closest distance, if there is data that is moved back to the process of determining the initial Centroid. Still, if there is none, it will display the clustering results.

2.19. System Design

The system will be designed using the Unified Modeling Language (UML) to describe its work process visually, making it easy to understand how it will work and help implement it into the program. Use case diagrams are drawn to identify actors or users and features in the system and explain what features can be implemented by each actor or user in the system.
The diagram above shows that the system will be designed as an actor with an admin role in the use case image. Users have full rights to access all features in the system, starting from login, home, the add cluster data menu, the check graph menu, and log out of the account. Class diagrams represent the structure and description of classes, attributes, methods, and relationships of each class used in the system. Activity diagrams define a sequence of activity flows or workflows carried out by the system, database, and each actor in the system. Sequence Diagrams display the behavior of objects in the Use Case by providing the period of existence of the object and the messages received or sent.
The relationship between the admin and all objects in the login process. For the first time, enter the admin username and password. Thus, the system will verify the username and password entered, and if the username and password are correct, the admin will be directed to the main page. But if wrong, the admin will be asked to re-enter the appropriate username and password—the relationship

between the admin and all objects in the home process. The first time the admin accesses the home, on this page, the admin can see student data and calculation results using the k-means method and can change and delete data. If the admin wants to change one of the data points, the admin needs to click the change button, and a form that can be changed will appear. After the admin has finished filling in the data form that he wants to change, click save, and then the changed data will be immediately saved in the database, and the data will be updated again, as well as in the data deletion process. The relationship between the admin and all objects when adding student data. The first time the admin accesses the add student data page. The admin can only add student data on this page by filling in the available data form. After the admin has finished filling in all the data, click save. Then the data that has just been added will be saved in the database. The relationship between the admin and all objects on the check graph page. Which admin will start by opening the check graph menu? This page only displays the graphs. The relationship between the admin and all objects in the process on the initial data page. Which admin will start by accessing the main page and selecting the initial data page? On this page, the admin will only see the initial data on the level of futsal interest. The relationship between the admin and all objects in the logout process. Which admin will start by accessing the logout menu? After the logout process, the admin is directed to the login menu. This system uses MySQL as a database management system to store data in a structured manner and provide convenience in data processing.

The author collected data in this study by distributing questionnaires to all Malikussaleh University students using Google Forms. The data were from 100 Malikussaleh University students who filled out the questionnaire. The questionnaire contained five questions about Futsal playing time, futsal field facilities, motivation to play Futsal, environmental factors, and plans in the field of Futsal. After all the required data was collected, the author then carried out the calculation process by applying the K-Means algorithm using the Manhattan distance formula, showing the results that cluster one had 20 data points, cluster 2 had 28 data points, cluster 3 had 5 data points, and cluster 4 had 47 data points.

2.20. Dataset

The dataset is a set of data the author uses to test the K-Means Clustering algorithm. The total data used is 100 Malikussaleh University student data.

3. Results and Discussion

3.1. Dataset

This data was obtained from students from several departments at Malikussaleh University as test material for the K-Means clustering of 100 students, as in Table 1.

Table 1. Data on Malikussaleh University students

No	Name	Department	Playtime	Field facilities	Play motivation	Environmental factors	Plans
1	Faris Bagaswara	Information Technology	8	4	4	3	2
2	Alvin Alvesaldy	Information Technology	5	2	3	2	1
3	RAFI ALFARISI	Information Technology	6	2	3	2	1
4	Taufik Hidayat	Communication Science	7	3	4	3	2
5	Hawari Muflih Munte	Information Technology	3	3	2	1	2
	.....		.....	.....	.....	.....	.....
96	fikri azzaki	Electrical Engineering	3	2	3	4	2
97	daudi bahari	Mechanical Engineering	2	2	1	2	1
98	win saradi	Law	2	2	1	2	1
99	muhammad fachri zidan	Accounting	11	3	4	2	2
100	Della Novita	Communication Science	0	1	1	2	1

3.1.1. Centroid test data

Determination of the initial cluster center as test data for four (4) centers is carried out randomly, as in Table 2.

Table 2. The initial Centroid of the test data

No	Name	Department	Playtime	Field facilities	Play motivation	Environmental factors	Plans
1	Faris Bagaswara	Information Technology	8	4	4	3	2
8	Syahrul Andika	Information Technology	4	2	1	2	1
13	Iqbal Armanda	Information Technology	8	4	4	2	2
21	Iqbal rizantha	Communication Science	2	2	1	2	1

Based on Table 2, the initial Centroid obtained for the data test shows that the five (5) variables have the highest value as the central cluster, while the smallest value is the lowest cluster.

3.2. Implementation of Min-Max Normalization

At this stage, the researcher performed calculations using min-max normalization. The Minimal-Maximal normalization was obtained with different distributions for each participant from 100 students, as in Table 3.

Table 3. Results of applying Min-Max normalization

No	Name	P1	P2	P3	P4	P5
1	Faris Bagaswara	0.727272	1	1	0.666666	1
2	Alvin Alvesaldy	0.454545	0.333333	0.666666	0.333333	0
3	RAFI ALFARISI	0.545454	0.333333	0.666666	0.333333	0
4	Taufik Hidayat	0.636363	0.666666	1	0.666666	1
5	Hawari Muflih Munte	0.272727	0.666666	1	0	1
	.....		.....	.....	.....	.....
96	Fikri azzaki	0.272727	0.333333	0.666666	1	1
97	Daudi Bahari	0.181818	0.333333	0	0.333333	0
98	win saradi	0.181818	0.333333	0	0.333333	0
99	Muhammadfachri zidan	1	0.666666	1	0.333333	1
100	Della Novita	0	0	0	0.333333	0

At this stage, calculations will be carried out using the K-Means Clustering algorithm to group student data from the Faculty of Engineering, Malikussaleh University. The input parameters used are datasets with an initial centroid of 4 data. The steps taken are as follows:

a. Determination of the initial cluster center (Centroid). This can be seen in Table 4.
b. Calculation using Manhattan Distance.

After the calculation, the closest distance and cluster group will be obtained for each level of futsal interest data for Malikussaleh University students, as seen in the table below.

Table 4. The closest distance and cluster data group for each data point

No	Name	C1	C2	C3	C4
1	Faris Bagaswara	0	3.363636364	0.333333333	3.545454545
2	Alvin Alvesaldy	2.606060606	0.757575758	2.272727273	0.939393939
3	RAFI ALFARISI	2.515151515	0.848484848	2,181818182	1.03030303
4	Taufik Hidayat	0.424242424	2.939393939	0.757575758	3.121212121
5	Hawari Muflih Munte	1.45454545	2.757575758	1.121212121	2.757575758

	.....	...	...	...	...
96	Fikri azzaki	1.787878788	2.424242424	2,121212121	2.424242424
97	Daudi Bahari	3.545454545	0.181818182	3.212121212	0
98	win saradi	3.545454545	0.181818182	3.212121212	0
99	Muhammadfachri zidan	0.939393939	2.969696979	0.606060606	3.151515152
100	Della Novita	4.060606061	0.696969697	3.727272727	0.515151515

After each data point, the distance for each cluster is calculated. The next step is to group the data according to its cluster. The cluster group of data is taken from the shortest distance of the data to a cluster. Grouping was carried out on 100 data points, as shown in Table 5.

Table 5. Cluster Grouping

No	Name	C1	C2	C3	C4	*Cluster*
1	Faris Bagaswara	0	3.363636364	0.33333333	3.545454545	C1
2	Alvin Alvesaldy	2.606060606	0.757575758	2.272727273	0.939393939	C2
3	RAFI ALFARISI	2.515151515	0.848484848	2,181818182	1.03030303	C2
4	Taufik Hidayat	0.424242424	2.939393939	0.757575758	3.121212121	C1
5	Hawari Muflih Munte	1.45454545	2.757575758	1.121212121	2.757575758	C3
	.....	...	...	...	...	
96	Fikri azzaki	1.787878788	2.424242424	2,121212121	2.424242424	C1
97	Daudi bahari	3.545454545	0.181818182	3.212121212	0	C4
98	win saradi	3.545454545	0.181818182	3.212121212	0	C4
99	Muhammadfachri zidan	0.939393939	2.969696979	0.606060606	3.151515152	C3
100	Della Novita	4.060606061	0.696969697	3.727272727	0.515151515	C4

After the new centroid value is obtained, the next step is to compare it with the previous centroid value. If the value is the same, the iteration process is stopped. However, the iteration process must be repeated if the centroid value differs. The results of the fourth iteration calculation show that the Centroid value is the same, and the calculation is a comparison to stop the clustering process.

3.3. Determining New Centroids

The determination of new clusters is based on the smallest distance, as shown in Table 6.

Table 6. New centroids

Centroid	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5
C1	0,601731602	0,666666667	0,936507937	0,761904762	1,047619
C2	0,421717172	0,490740741	0,407407407	0,509259259	0
C3	0,6	0,733333333	0,866666667	0,2	1
C4	0,129186603	0,429824561	0,043859649	0,333333333	0

Description:
Variable 1 = Playing Time; Variable 2 = Field Facilities; Variable 3 = Playing Motivation
Variable 4 = Environmental Factors; Variable 5 = Future Plans

3.4. Comparison of Cluster Grouping Results

Table 7. Comparison of cluster grouping

Cluster	Number
professional player	20

regular player	28
Beginner player	5
amateur player	47
Total	100

The comparison of all futsal players' interests is shown in Figure 3, which is the result of the clustering calculation of futsal interest levels among Malikussaleh University students.

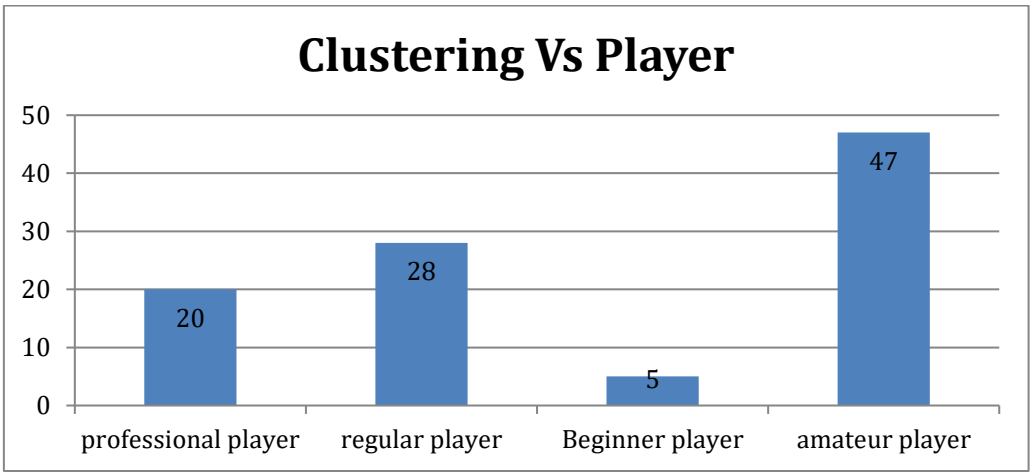

Fig 3. Comparison of all futsal enthusiasts at Malikussaleh University.

4. Conclusion

Data clustered into professional players is 20%, regular players 28%, beginner players 5%, and amateur players 47%. So it can be concluded that amateur players are pretty high, followed by regular players, and professional players, and beginner players are very few, so it is necessary to provide coaching to Malikussaleh University students who are directed to become professional players to raise the name of the University in the Futsal player event in the future.

Acknowledgment

The researcher would like to thank the Chancellor and staff, the Dean and staff, and the supervising lecturer who has provided direction in completing this research. Hopefully, it will be helpful to amend.

References

[1] Nandang Iriadi, Lutfi Setioningtias, Priatno Priatno., "Implementasi Data Mining Pada Klasifikasi Ketidakhadiran Pegawai Menggunakan Metode C4. 5", *Comput. Sci.*, vol. 1, nr. 25, pp. 53–61, 2021.
[2] S. Seimahuira, "Berdasarkan Online Reviews Tripadvisor Menggunakan", vol. 12, nr. 1, pp. 53–58, 2021.
[3] James C Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1981.
[4] James C Bezdek, Robert Ehrlich, William Full., "FCM: The fuzzy c-means clustering algorithm", *Comput. Geosci.*, vol. 10, nr. 2–3, pp. 191–203, 1084, [Online]. Beschikbaar op: https://doi.org/10.1016/0098-3004(84)90020-7
[5] A Muh. Ismail, "Analisis Minat pada Kegiatan Ekstrakurikuler Futsal SNA Negeri 8 Maros", in *Skripsi*, Makasar, 2020, p. 18. [Online]. Beschikbaar op: http://eprints.unm.ac.id/id/eprint/18133
[6] Atang Supriatna, William Dharmawan, Christina Juliane., "Algoritma K-Means Clustering Pada Pengelompokan Minat Bakat Siswa SMK PGRI 2 Karawang", *J. Tek. Inform. Dan Sist. Inf.*, vol. 10, nr. 1, pp. 38–50, 2023.
[7] Mardiansa, Herlina Latifa Sari, Prahasti., "Penerapan Data Mining Untuk Mengetahui Minat Siswa Pada Pelajaran IPA Mengunakan Metode K-Means Clustering", *Multi Disiplin Dehasen (MUDE)*, vol. 2, nr. 4, pp. 693–702, 2023.
[8] Maria Dolores Reina Paz and Juan Carlos Rodriguez Vargas, "Main theoretical consumer behavioral models. A review from 1935 to 2021", *Heliyon*, vol. 9, nr. 3, p. 51, 2023, [Online]. Beschikbaar op: https://doi.org/10.1016/j.heliyon.2023.e13895
[9] Ryan Salviany Grova, Mohammad Zaky. Ahmad Hamidi., "KORELASI ANTARA KEPERCAYAAN DIRI DAN PERFORMA PUNCAK PADA ATLET FUTSAL PROFESIONAL", *Jurnla Olahraga dan Kesehat. Indones.*, vol. 4, nr. 2, pp. 148–158, 2024, [Online]. Beschikbaar op: https://jurnal.stokbinaguna.ac.id/index.php/JOK
[10] A. K. Wahyudi, N. Azizah, en H. Saputro, "Data Mining Klasifikasi Kepribadian Siswa Smp Negeri 5 Jepara Menggunakan Metode Decision Tree Algoritma C4.5", *J. Inf. Syst. Comput.*, vol. 2, nr. 2, pp. 8–13, 2022.
[11] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, en R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang", *J. Janitra Inform. dan Sist. Inf.*, vol. 2, nr. 1, pp. 9–19, 2022, doi:

10.25008/janitra.v2i1.142.
[12] Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Guido Caldarelli, Paolo Cintia, Stefano Cresci, Angelo Facchini, Fosca Giannotti, Aristides Gionis, Riccardo Guidotti, Michael Mathioudakis, Cristina Ioana Muntean, Luca Pappalardo, Dino Pedreschi, Ev, "Big Data and the transformation of the city", *Int. J. Data Sci. Anal.*, vol. 11, pp. 311–340, 2021.
[13] Jaap Beltman, Marcos R Machado, Joerg R Ostrerrieder., "Predicting retail customers' distress in the finance industry: An early warning system approach", *J. Retail. Consum. Serv.*, vol. 82, pp. 1–17, 2025, [Online]. Beschikbaar op: https://doi.org/10.1016/j.jretconser.2024.104101
[14] Jose Alejandro Cano, Abraham Allec Londono-Pineda, Emiro Antonio Campo, Sergio August Fernandez., "Sustainable business models of e-marketplaces: An analysis from the consumer perspective", *J. Open Innov. Technol. Mark. Complex.*, vol. 9, nr. 3, pp. 1–18, 2023, [Online]. Beschikbaar op: https://doi.org/10.1016/j.joitmc.2023.100121
[15] Kamal Malik, Harsh Sadawarti, Gursharanjeet Singh Kalra., "Comparative Analysis of Outlier Detection Techniques", *Int. J. Comput. Appl.*, vol. 97, nr. 8, pp. 12–21, 2014, [Online]. Beschikbaar op: http://dx.doi.org/10.5120/17026-7318
[16] R. Kurniawan en R. Dewi, "Penerapan Algoritma K-Means Clustering Dalam Persentase Merokok Pada Penduduk Umur Di Atas 15 Tahun Menurut Provinsi", *J. Sist. Komput. Dan Inform. Hal*, vol. 2, nr. 2, pp. 178–186, 2021, doi: 10.30865/json.v2i2.2770.
[17] Caroline X. Gao, Dominic Dwyer, Ye Zhu, Catherine L. Smith, Lan Du, Kate M. Filia, Johanna Bayer, Jana M. Menssink, Teresa Wang, Christoph Bergmeir, Stephen Wood, Sue M. Cotton., "An overview of clustering methods with guidelines for application in mental health research", *Psychiatry Res.*, vol. 327, pp. 1–28, 2023, [Online]. Beschikbaar op: https://doi.org/10.1016/j.psychres.2023.115265
[18] Siegfried M Rump, "Verified bounds for the determinant of real or complex point or interval matrices", *J. Comput. Appl. Math.*, vol. 372, pp. 1–10, 2020, [Online]. Beschikbaar op: https://doi.org/10.1016/j.cam.2019.112610
[19] Paula Kivimaa, Wouter Boon, Sampsa Hyysalo, Laurens Klerkx., "Towards a typology of intermediaries in sustainability transitions: A systematic review and a research agenda", *Res. Policy*, vol. 48, nr. 4, pp. 1062–1075, 2019, [Online]. Beschikbaar op: https://doi.org/10.1016/j.respol.2018.10.006
[20] Angeli Dwiyanti Nur'azizah and Zaehol Fatah, "Implementasi Algoritma K-Means Clustering Untuk Perokok Usia Di Atas 15 Tahun", *Hudang J. Multidsiplin Ilmu*, vol. 2, nr. 10, pp. 133–139, 2024.
[21] L. Fimawahib en E. Rouza, "Penerapan K-Means Clustering pada Penentuan Jenis Pembelajaran di Universitas Pasir Pengaraian", *INOVTEK Polbeng - Seri Inform.*, vol. 6, nr. 2, p. 234, 2021, doi: 10.35314/isi.v6i2.2096.
[22] G. S. Nugraha en H. Hairani, "Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode K-Means", *J. MATRIK*, vol. 17, nr. 2, pp. 13–23, 2018, doi: 10.30812/matrik.v17i2.84.
[23] Amri Muliawan Nur,Muhammad Saiful, Hariman Bahtiar, Muhammad Taufik Hidayat., "Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi", *infotek J. Inform. dan Teknol.*, vol. 7, nr. 2, pp. 478–488, 2024.
[24] D. A. Nasution, H. H. Khotimah, en N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN", *Comput. Eng. Sci. Syst. J.*, vol. 4, nr. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
[25] J. A. S. Siregar en K. Handoko, "pengembangan sistem presensi karyawan dengan teknologi GPS berbasis web", *J. Comasie*, vol. 6, nr. 2, p. 3, 2021.
[26] A. A. Rizal *e.a.*, "Peningkatan Efektifitas Programming Dengan Pelatihan Python for Data Science Bagi Komunitas Programming Pondok Pesantren", *J. Pengabdi. Masy.*, vol. 1, nr. 1, pp. 13–19, 2021.
[27] Deni Arsandi, Novita br Ginting, Dewi Primasari., "Web-Based Lecturer Performance Selection Information System Using Weighted Product (WP) Method", *J. Mantik*, vol. 5, nr. 2, pp. 615–619, 2021, [Online]. Beschikbaar op: http://www.iocscience.org/ejournal/index.php/mantik/index
[28] B Budiman, "Perbandingan Algoritma dan Penerapannya", *Nuansa Inform.*, vol. 15, nr. 2, pp. 37–52, 2021.
[29] H. W. Pramana, "Rancang Bangun Aplikasi Fitness Berbasis Android (Studi Kasus : Popeye Gym Suwaan)", *E-journal Tek. Inform.*, vol. 1, nr. 2, pp. 1–10, 2020.
[30] Rizky Dwi Kurniawan, Yudie Irawan, Arif Setiawan, Noor Latifah., "Perancangan Sistem Informasi Pelayanan Pada Laboratorium Klinik Prima Medika Kudus Berbasis Web Menggunakan UML (Unified Modeling Language)", *J. Sist. Inform. dan Teknol.*, vol. Edisi Khus, pp. 27–34, 2024.
[31] Mohamad Sugeng Pangestu dan Maulida Ayu Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means", *Sainteks*, vol. 19, nr. 2, pp. 141–155, 2022.