Application of Data Mining on Player Statistics for Scouting in Football

Gandung Triyono[*1], Aditya Agus Wisanto[2], Achmad Fachrurozy[3]
[1,2,3]Master of Computer Science Informatics Engineering, Budi Luhur University, Indonesia
E-mail: [*1]**gandung.triyono@budiluhur.ac.id**, [2]2311600015@student.budiluhur.ac.id,
[3]2311600395@student.budiluhur.ac.id

Abstract

The sophistication of today's technology makes the use of data increasingly massive. All digital aspects must have data that is ready to be processed, including in the football industry. The use of data in the football industry is one of them used to record all activities carried out by players to see their performance in the match. Dewa United has a scouting division that is tasked with finding talented players according to the wishes of the head coach. In its search, the scouting division observes the players on the field and also uses raw statistical data to see the player's performance. However, the implementation of these activities still has obstacles as evidenced by the difference between the results of observations and the performance of players when joining the team. To solve this problem, the use of data mining can provide scouting recommendations according to player statistics, making the scouting process effective and efficient. The purpose of this study is to make it easier for the team to search for players according to what is desired, which is obtained is a web-based application that has a scouting recommendation feature based on attributes or players according to choice and detailed descriptions of the selected players..

Keywords *— Scouting, Football, Data Mining*

1. INTRODUCTION

Football is one of the sports that is very popular with all groups, both young and old. The simplicity of the game of football makes football one of the sports that is popular with many parties. Each team that competes has the main task of scoring as many goals as possible and avoiding conceding goals. This ease and simplicity are what make football playable and popular with all parties.

The main task of football is to score goals and prevent conceding goals, which is used by coaches and analysts in the team to create strategies. The strategies created can vary depending on the quality and condition of each team. There is no perfect strategy in football so coaches and analysts are always looking for and experimenting to make their teams always win.

Football, which was originally just a game of processing the ball, has now become increasingly complex with data mining. Each player and tactical instructions from the coach can be used as data that can be analyzed into valuable information. One example is football event data, football event data is all the activities that occur during a football match. Passing,

shooting, tackles are some of the activities in football event data. With this football event data, all players have their own characteristics that are described through data.

Dewa United Football Club has eight staff to help players achieve their best game. Each includes a head coach, goalkeeping coach, two assistant coaches, a doctor, a physiotherapist, a translator, and a video analyst. Dewa United also has a scouting division to find talented players that are used to recruit players into the team.

One way to make it easier to find players according to their game is to apply reference data in viewing players. Coaches and analysts who initially have subjective assessments can be helped by objective data.

2. RESEARCH METHOD

2.1. System Analysis Method

To identify a problem and make strategic plans, it is necessary to analyze the procedures that are running. The following are metrics to identify the plans needed to find the best players through scouting.

Tabel 1. *Critical Success Factor*

Critical Success Factor	***Weight***	***Rating***	***Score***
Observing players directly	0.2	3	0.6
Taking advantage of player statistical data	0.3	2	0.6
Offer from agent	0.1	3	0.3
Live discussion between scout and coach division	0.4	3	1.2
TOTAL	1.0	11	2.7

Based on the table above, Direct discussion between the scout and coach divisions has the highest weight followed by data utilization. This makes discussion and data utilization important.

2.2. Input Analysis, Process Analysis, Output Analysis

1) Input Analysis
   Input analysis contains the data needed to be processed into a decision in the scouting process.
   a. Input Name : Targeted player data
      Function : As a player recommendation from the scouting division
      Source : Observation
      Description : Player name, age, value price, position, attributes

b. Input Name : Player data from agent
Function : As a player recommendation from agent
Source : Agent
Description : Player name, age, value price, position, attributes

2) Process Analysis

Explains the process that occurs that is needed to discuss a problem.

Module Name: Player discussion
Input: Player data from the scouting and agent divisions
Output: Final player report
Process Summary: The scouting division and head coach hold discussions to obtain final results regarding the players.

3) Output Analysis

Provides information regarding the output of the process.

Output Name: Final player report
Function: As a recommendation for players from the scouting division and head coach to be given to the club president
Media: PDF and Paper
Distribution: Scouting division

2.3. Problems Faced and Alternative Problem Solving

1) Problems Faced

Based on the results of observations and interviews that have been conducted, there are several problems for Dewa United Football Club in finding talented players or scouting. The problems are described as follows: 1. Differences in observation results and player performance when joining the team, 2. Use of data that is only based on raw data via Wyscout, 3. The absence of an internal club system to provide convenience in finding and analyzing targeted players.

2) Alternative Problem Solving

The problems in the points above can be minimized or resolved by creating a new web-based system that can: 1. Display player statistical performance so that you can find out the events that players do, 2. Display data and recommendations using machine learning so that you can produce more specific and accurate decisions.

Conducting a search for targeted players and providing specific details regarding the player's value price, age, etc. so that it is easy. To make it easier to know and see the procedures for the scouting process at Dewa United Football Club, this study uses Unified Modeling Language (UML) to describe the procedures that are running.

2.4. Scouting System Analysis on Use Case Diagram

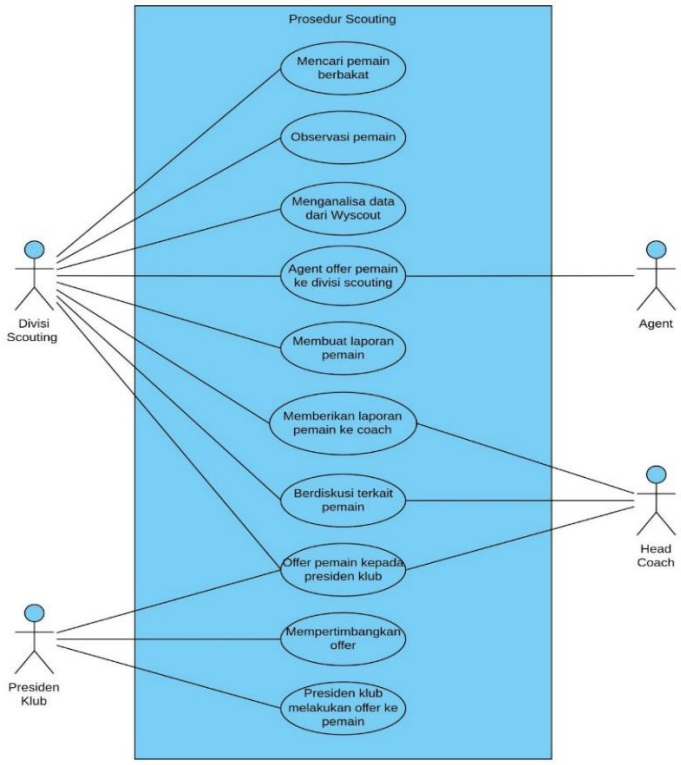


Figure 1. Use Case Diagram of the Running System

Based on the Use Case Diagram above, there are One use case diagram system that covers all scouting procedures and Four actors who carry out activities and ten use cases carried out by the actors include: a. The scouting division that carries out activities to find talented players, observe players, analyze data from Wyscout, receive offers from agents, make player reports, provide player reports to the head coach, discuss players, and make player offers to the club president, b. Agents who carry out activities to offer players to the scouting division, c. Head coaches who carry out activities to receive player reports, discuss players, and make player offers to the club president, d. Club presidents who carry out activities to receive player offers from the scouting division and head coach, consider offers, and make offers to players.

3. RESEARCH RESULTS AND DISCUSSION

Based on the problem, the application of scouting is still traditional by utilizing observations and also offers from agents. Although it has implemented data as additional analysis, the data used is still raw so that subjective potential can still apply. This makes the scouting process not on target. As a proposed improvement, a web-based system is needed that contains statistical data from all players competing in League 1 2023/2024. The statistical data is processed to provide player recommendations that match the selected players. Recommendations for players that have been received will be considered by the club president for recruitment.

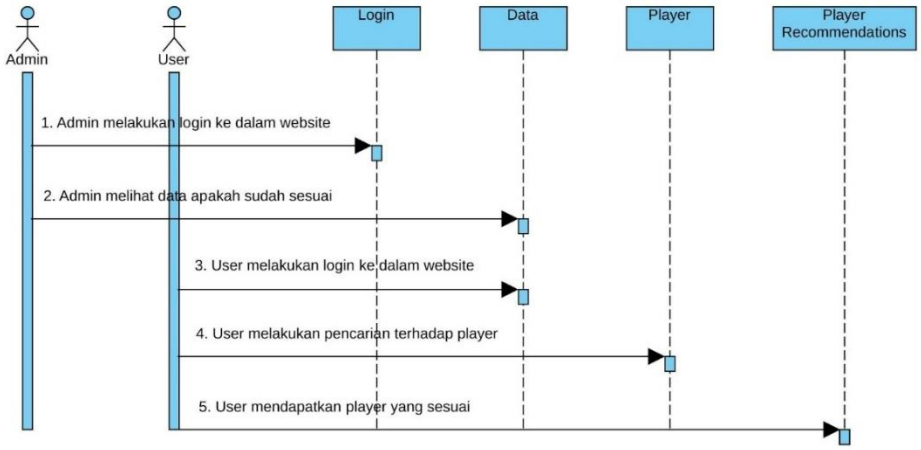


Figure 2. Sequence Diagram of the Proposed System

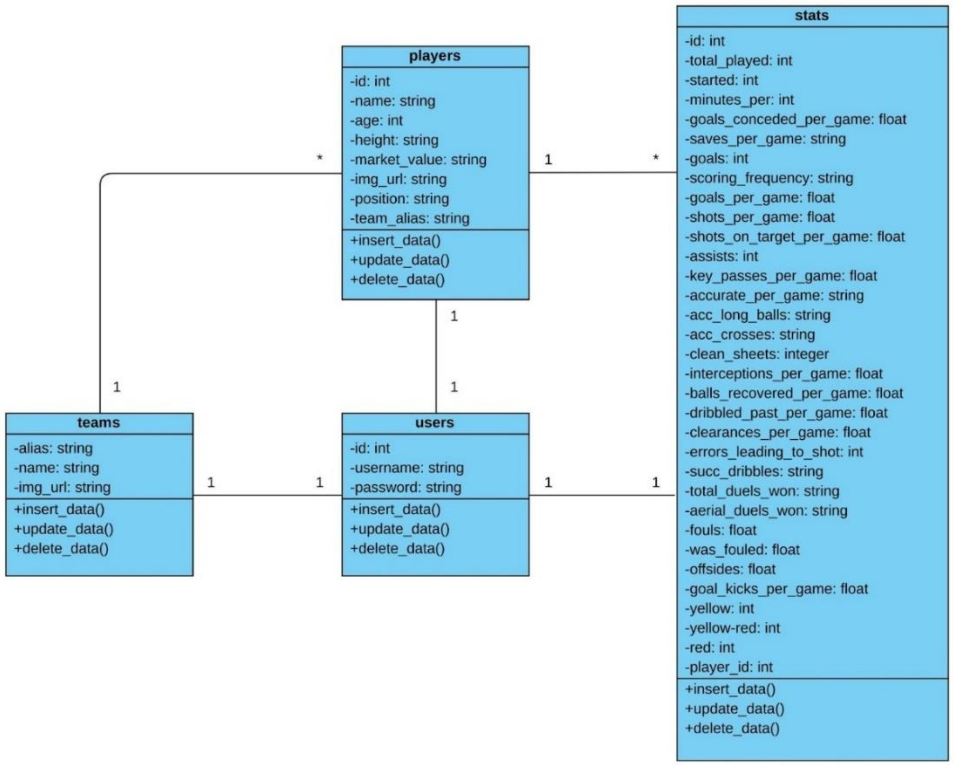


Figure 3. Class Diagram of the Proposed System

3.1. Data Mining Process

In accordance with the title of this study, namely the application of data mining to player statistics, there is a process of taking, analyzing and processing data into information that can be consumed. The data used here is data taken from all players competing in the 2023/2024 Indonesian League 1 season which is taken from the Sofascore application.

3.2. Data Processing

The following is the head data taken from the sofascore application.

[3]:

	Name	Age	Height	Market Value	img_url	Position	Team	Total played	Started	Minutes per game	...	Succ. dribbles	Total duels won	Aerial duels won	Fouls	Was fouled	Offsides	Goal kicks per game	Yellow
0	Alex	31	1,87 m	Rp4,35B	https://lapangbola-rails-files.s3.ap-southeast...	CF	Dewa United FC	32	30	94	...	1.9 (70%)	7.0 (37%)	1.7 (50%)	0.9	1.7	0.6	NaN	3
1	Majed Osman	30	1,83 m	Rp6,52B	https://lapangbola-rails-files.s3.ap-southeast...	LW	Dewa United FC	30	21	70	...	4.9 (75%)	8.8 (47%)	0.2 (33%)	1.0	1.4	0.3	NaN	3
2	Septian Bagaskara	26	1,83 m	Rp2,61B	https://lapangbola-rails-files.s3.ap-southeast...	CF	Dewa United FC	19	4	29	...	0.3 (55%)	1.8 (44%)	1.1 (61%)	0.3	0.2	0.1	NaN	3
3	Rangga Muslim Perkasa	30	1,66 m	Rp2,17B	https://lapangbola-rails-files.s3.ap-southeast...	AMF	Dewa United FC	9	5	43	...	0.8 (88%)	2.4 (32%)	0.2 (33%)	1.0	0.8	0.1	NaN	2
4	Egy Maulana Vikri	24	1,70 m	Rp4,35B	https://lapangbola-rails-files.s3.ap-southeast...	RW	Dewa United FC	29	25	81	...	3.2 (76%)	6.6 (40%)	0.3 (24%)	1.2	1.3	0.2	NaN	1

5 rows × 38 columns

Figure 4. Head Data

Apart from that, you need to know how much data it contains and what columns the data contains.

```
[4]: df.shape

[4]: (458, 38)

[5]: df.columns

[5]: Index(['Name', 'Age', 'Height', 'Market Value', 'img_url', 'Position', 'Team',
       'Total played', 'Started', 'Minutes per game',
       'Goals conceded per game', 'Saves per game', 'Goals',
       'Scoring frequency', 'Goals per game', 'Shots per game',
       'Shots on target per game', 'Assists', 'Key passes per game',
       'Accurate per game', 'Acc. long balls', 'Acc. crosses', 'Clean sheets',
       'Interceptions per game', 'Balls recovered per game',
       'Dribbled past per game', 'Clearances per game',
       'Errors leading to shot', 'Succ. dribbles', 'Total duels won',
       'Aerial duels won', 'Fouls', 'Was fouled', 'Offsides',
       'Goal kicks per game', 'Yellow', 'Yellow-red', 'Red'],
      dtype='object')
```

Figure 5. Check Data

It can be seen above that there are 458 data and many columns that decorate the data. Before being processed, the data needs to be checked for missing values that will cause the data to be difficult to process.

```
Name                          0
Age                           0
Height                        0
Market Value                  0
img_url                       0
Position                      0
Team                          0
Total played                  0
Started                       0
Minutes per game              0
Goals conceded per game     413
Saves per game              413
Goals                         0
Scoring frequency             0
Goals per game                0
Shots per game                0
Shots on target per game      0
Assists                       0
Key passes per game           0
Accurate per game             0
Acc. long balls               0
Acc. crosses                  0
Clean sheets                418
Interceptions per game        0
Balls recovered per game      0
Dribbled past per game        0
Clearances per game           0
Errors leading to shot        0
Succ. dribbles                0
Total duels won               0
Aerial duels won              0
Fouls                         0
Was fouled                    0
Offsides                      0
Goal kicks per game         394
Yellow                        0
Yellow-red                    0
Red                           0
```

Figure 6. Missing Values

There are many missing values in the Goals conceded per game, Saves per game, Clean sheets, Goal kicks per game columns. If you look at the data above, the data is the goalkeeper attribute. Therefore, the empty data only needs to be filled with the number 0.

```
df.fillna(0, inplace = True)
```

Figure 7. Input Missing Values

In the Saves per game, Scoring frequency, Accurate per game, Acc. long balls, Acc. crosses, Succ. dribbles, Total duels won, Aerial duels won columns there is inconsistent data because it displays two different numbers.

	Saves per game	Scoring frequency	Accurate per game	Acc. long balls	Acc. crosses	Succ. dribbles	Total duels won	Aerial duels won
0	0	151 min	11.7 (77%)	0.4 (57%)	0.1 (25%)	1.9 (70%)	7.0 (37%)	1.7 (50%)
1	0	523 min	19.7 (81%)	0.6 (52%)	0.7 (33%)	4.9 (75%)	8.8 (47%)	0.2 (33%)
2	0	275 min	4.1 (84%)	0 (0%)	0.0 (0%)	0.3 (55%)	1.8 (44%)	1.1 (61%)
3	0	0 min	14.0 (89%)	0.6 (100%)	0.1 (25%)	0.8 (88%)	2.4 (32%)	0.2 (33%)
4	0	337 min	18.8 (78%)	0.7 (43%)	0.4 (27%)	3.2 (76%)	6.6 (40%)	0.3 (24%)
...	...	...	...	...	...	...	...	...
453	0	0 min	20.6 (80%)	1.8 (52%)	0.8 (50%)	0.5 (92%)	3.8 (32%)	1.3 (60%)
454	0	257 min	3.9 (78%)	0.1 (50%)	0.2 (50%)	0.4 (100%)	1.8 (36%)	0.6 (45%)
455	0	0 min	11.4 (84%)	0.6 (37%)	0.2 (100%)	0.1 (100%)	1.8 (32%)	0.4 (44%)
456	0	0 min	18.0 (84%)	0.5 (25%)	1.0 (67%)	0.0 (0%)	3.0 (35%)	1.0 (50%)
457	3.3 (77%)	0 min	12.8 (88%)	2.7 (66%)	0.0 (0%)	0.1 (100%)	1.1 (74%)	0.8 (83%)

458 rows × 8 columns

Figure 8. Inconsistent Data

The data needs to be converted into data that can be processed.

```
def process_value(x):
    x = str(x)
    if '(' in x and '%' in x:
        return float(x.split(' ')[0])
    else:
        return float(x)

df['Saves per game'] = df['Saves per game'].apply(lambda x: process_value(x))
df['Scoring frequency'] = df['Scoring frequency'].apply(lambda x: int(x.split('m')[0]))
df['Accurate per game'] = df['Accurate per game'].apply(lambda x: int(x.split('(')[1].split('%')[0]))
df['Acc. long balls'] = df['Acc. long balls'].apply(lambda x: int(x.split('(')[1].split('%')[0]))
df['Acc. crosses'] = df['Acc. crosses'].apply(lambda x: int(x.split('(')[1].split('%')[0]))
df['Succ. dribbles'] = df['Succ. dribbles'].apply(lambda x: float(x.split('(')[0]))
df['Total duels won'] = df['Total duels won'].apply(lambda x: int(x.split('(')[1].split('%')[0]))
df['Aerial duels won'] = df['Aerial duels won'].apply(lambda x: int(x.split('(')[1].split('%')[0]))
```

Figure 9. Correct inconsistent data

To facilitate data processing, a new column is required that contains general position data and also the ID of each player.

```
position_mapping = {
    'CF': 'Forward',
    'SS': 'Forward',
    'LW': 'Forward',
    'RW': 'Forward',
    'AMF': 'Midfielder',
    'CMF': 'Midfielder',
    'DMF': 'Midfielder',
    'LM': 'Midfielder',
    'RM': 'Midfielder',
    'CB': 'Defender',
    'LB': 'Defender',
    'RB': 'Defender',
    'GK': 'Goalkeeper'}

df['Position Group'] = df['Position'].map(position_mapping)
```

```
df['ID'] = df.index
```

Figure 10. Update Data

The following are the final results of the data that will be processed next.

[15]:

ıme	Age	Height	Market Value	img_url	Position	Team	Total played	Started	Minutes per game	...	Aerial duels won	Fouls	Was fouled	Offsides	Goal kicks per game	Yellow	Yellow-red	Red	Position Group	ID
\lex	31	1,87 m	Rp4,35B	https://lapangbola-rails-files.s3.ap-southeast...	CF	Dewa United FC	32	30	94	...	50	0.9	1.7	0.6	0.0	3	1	1	Forward	0
ajed nan	30	1,83 m	Rp6,52B	https://lapangbola-rails-files.s3.ap-southeast...	LW	Dewa United FC	30	21	70	...	33	1.0	1.4	0.3	0.0	3	0	0	Forward	1
tian cara	26	1,83 m	Rp2,61B	https://lapangbola-rails-files.s3.ap-southeast...	CF	Dewa United FC	19	4	29	...	61	0.3	0.2	0.1	0.0	3	0	0	Forward	2
gga slim asa	30	1,66 m	Rp2,17B	https://lapangbola-rails-files.s3.ap-southeast...	AMF	Dewa United FC	9	5	43	...	33	1.0	0.8	0.1	0.0	2	0	0	Midfielder	3
Egy ana /ikri	24	1,70 m	Rp4,35B	https://lapangbola-rails-files.s3.ap-southeast...	RW	Dewa United FC	29	25	81	...	24	1.2	1.3	0.2	0.0	1	1	1	Forward	4

Figure 11. Final Data Processing

3.3. Exploratory Data Analysis

To understand the data more completely, data visualization can be used in this study.

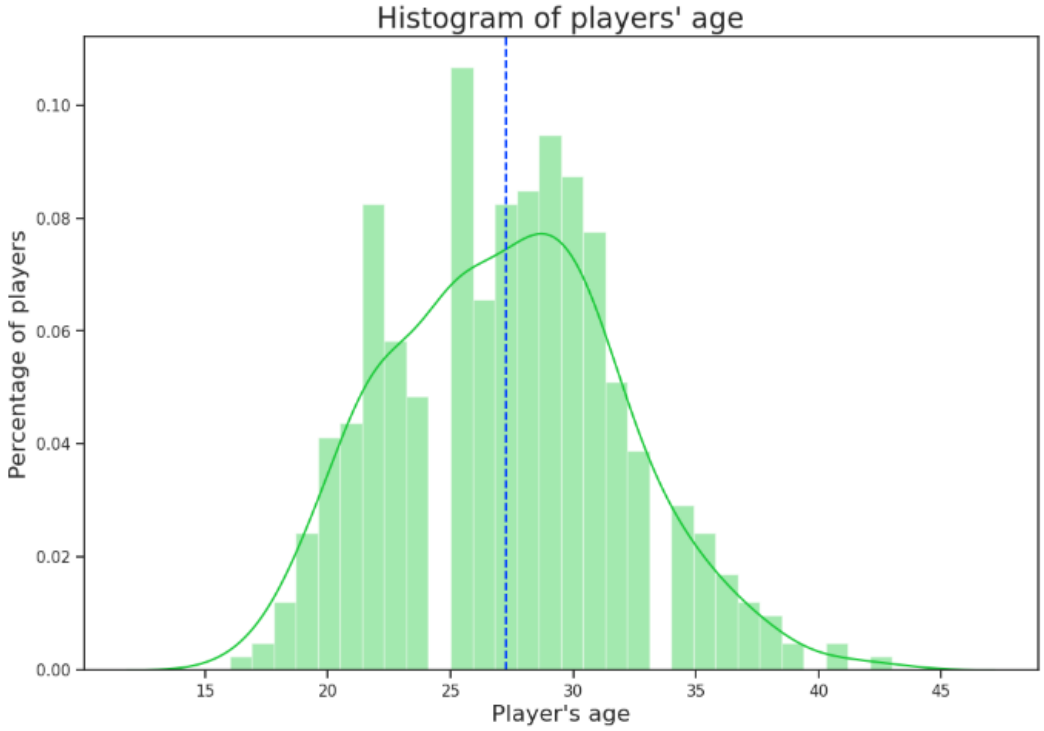


Figure 12. Age Histogram

According to the data above, it can be seen that the average age of the players is 27. Below is an analysis of the player position data.

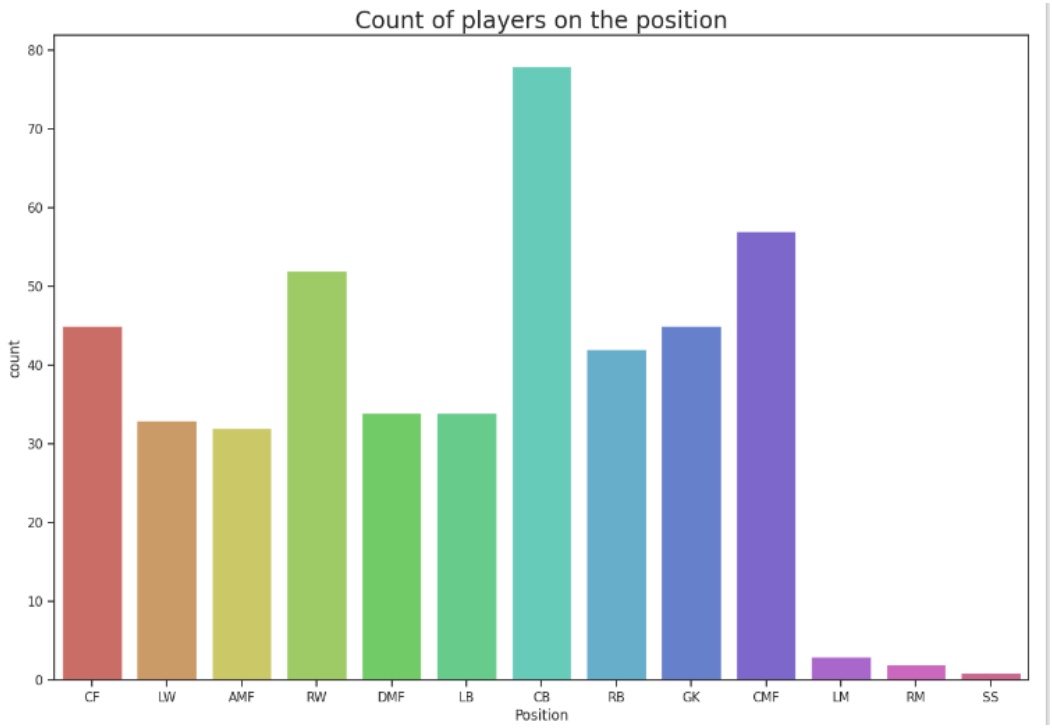


Figure 13. Position Count

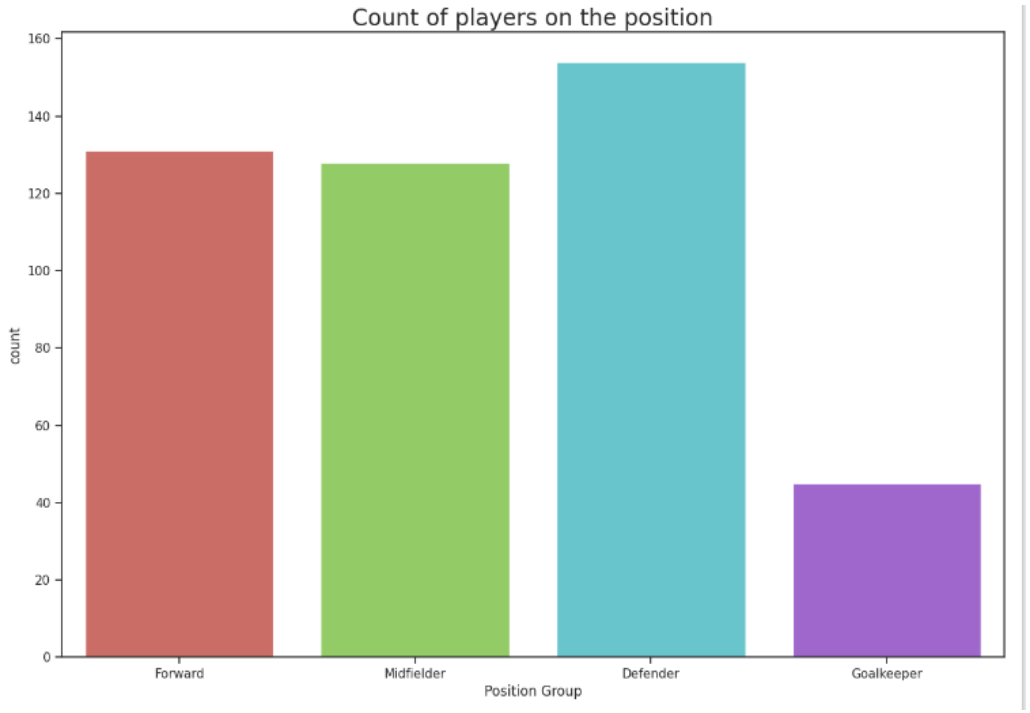


Figure 14. Position Group Count

The image above is the number of positions calculated from the general position and also the group position. To see the correlation of each variable, a heatmap of the data is needed.

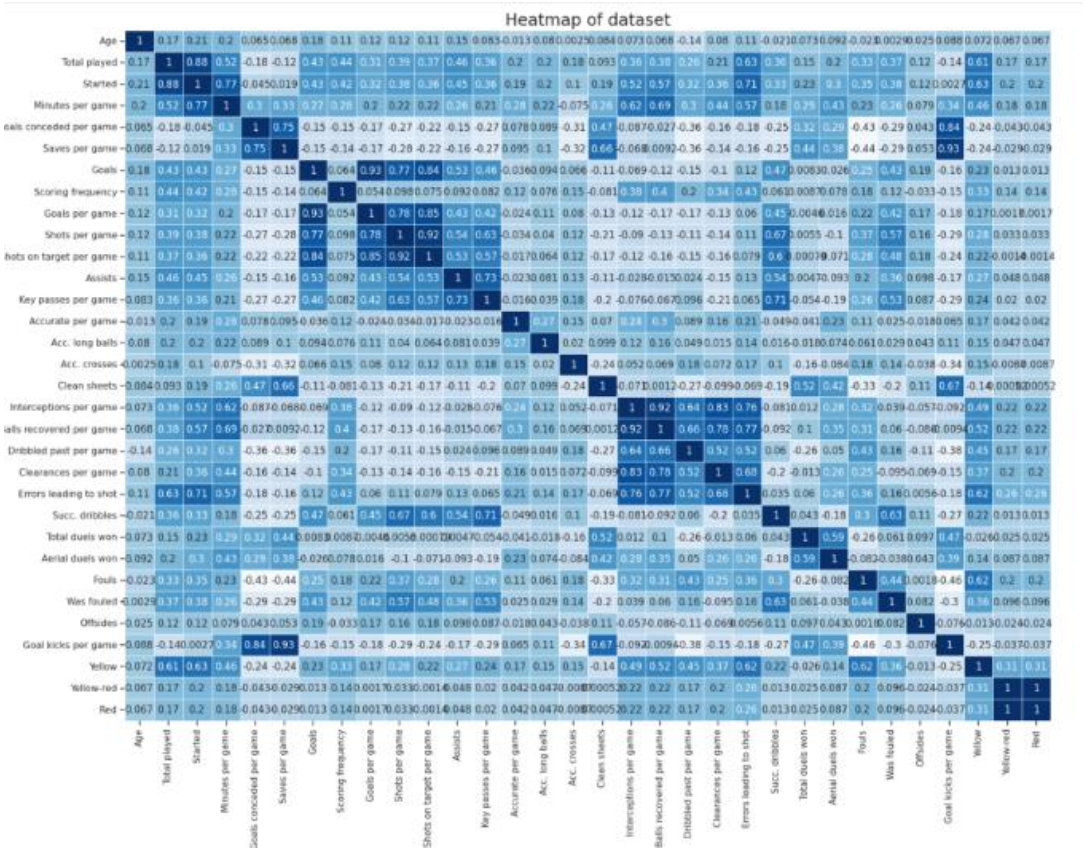


Figure 15. Heatmap of Dataset

It can be seen above that Total played, Started, and Minutes per game have a good correlation. Likewise with the attacking side such as variables Goals, Scoring frequency, Goals per game, Shots per game, Shots on target per game, Assists, and Key passes per game. On the defending side there is a good correlation between the variables Interceptions per game, Balls recovered per game, Dribbled past per game, Clearances per game, and Errors leading to shot.

3.4. Model Engineering

After the data can be understood, the data is processed into information that can be used in the future. In this study, the processed data is used as a model to provide information about player recommendations according to the data above. This study uses a machine learning algorithm model, namely K-Nearest Neighbors (KNN) to get player recommendations. The way KNN works is by placing data at a point and then looking for the closest point to be used as a recommendation. Here is a preview of the KNN code.

```
def recommended_k_players_df(player, k_players = 100):
    pos_group = list(df['Position Group'][df['Name']==player])[0]

    if pos_group == 'Forward':
        indices = list(df[(df['Position Group'] == 'Forward')|(df['Position Group'] == 'Midfielder')].index.values)
    elif pos_group == 'Midfielder':
        indices = list(df[(df['Position Group'] == 'Forward')|(df['Position Group'] == 'Midfielder')].index.values)
    elif pos_group == 'Defender':
        indicies = list(df[(df['Position Group'] == 'Defender')].index.values)
    elif pos_group == 'Goalkeeper':
        indices = list(df[(df['Position Group'] == 'Goalkeeper')].index.values)

    predictors_scaled_subset = predictors_scaled.iloc[indices, :]
    predictors_subset = predictors.iloc[indices, :]

    #Fit KNN for the k_players within that position group
    recommendations = NearestNeighbors(n_neighbors=k_players, algorithm='auto').fit(predictors_scaled_subset)

    #Pass the player name from the dataset to the function and get 5 similar players as output
    player_indices = recommendations.kneighbors(predictors_scaled_subset)[1]

    #Get player index
    df_subset = df.iloc[indices,:].reset_index()
    index = df_subset[df_subset['Name']==player].index.tolist()[0]

    #Make variables global
    global recommend_list
    global recommended_df
    global recommended_names

    recommended_list = []

    recommended_df = predictors_subset.iloc[list(player_indices[index][:]),:]
    recommend_list = list(df_subset.iloc[list(player_indices[index][:]),:].ID)
    recommend_list.insert(0, index)

    recommended_names = df['Name'][df.ID.isin(recommend_list)]
    return recommend_list, recommended_names, recommended_df
```

Figure 16. KNN Process

The following is one of the results obtained from the KNN process above.

```
player = str(input("Enter the player you are looking for: \n"))
num_of_players = input("Enter the number of similar players you are looking for: \n")

recommendation = recommended_k_players_df(player, (int(num_of_players)))
print("Here are {0} players similar to ".format(int(num_of_players)) +  player + ': \n \n' + str(recommendation[1][:]))
print("\n Here are the players' features ")
display(recommendation[2].set_index(pd.Index(df.Name[list(recommendation[0][1:])].values)))
```

```
Enter the player you are looking for:
 Egy Maulana Vikri
Enter the number of similar players you are looking for:
 4
Here are 4 players similar to Egy Maulana Vikri:

4      Egy Maulana Vikri
107          Renan Silva
204          Ricky Cawor
258         Ariel Lucero
Name: Name, dtype: object

 Here are the players' features
```

	Age	Total played	Started	Minutes per game	Goals conceded per game	Saves per game	Goals	Scoring frequency	Goals per game	Shots per game	...	Succ. dribbles	Total duels won	Aerial duels won	Fouls	Was fouled	Offsides	Goal kicks per game	Yellow
Egy Maulana Vikri	24	29	25	81	0.0	0.0	7	337	0.2	2.1	...	3.2	40	24	1.2	1.3	0.20	0.0	1
Renan Silva	35	28	28	88	0.0	0.0	6	409	0.2	3.5	...	2.9	48	26	1.1	2.0	0.04	0.0	6
Ricky Cawor	26	25	16	56	0.0	0.0	7	199	0.3	1.4	...	1.8	42	40	1.0	1.2	0.30	0.0	4
Ariel Lucero	25	26	22	80	0.0	0.0	0	0	0.0	1.6	...	2.3	46	27	0.9	1.3	0.08	0.0	4

4 rows × 32 columns

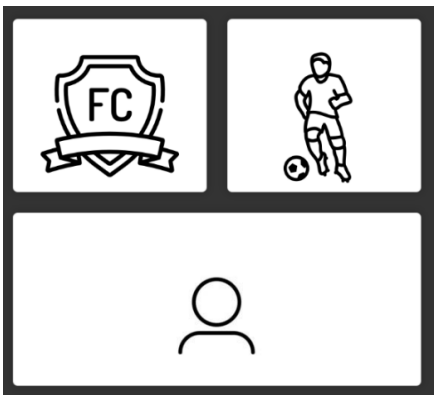
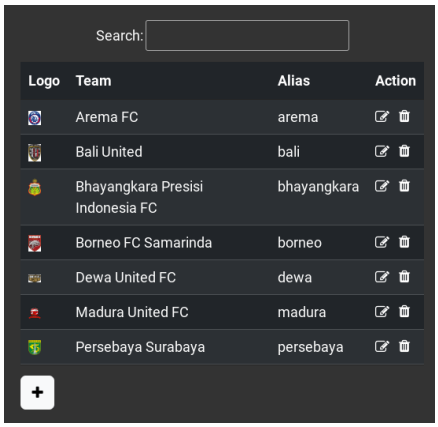
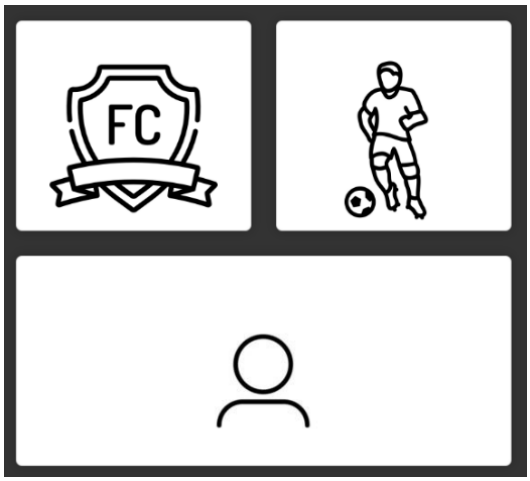
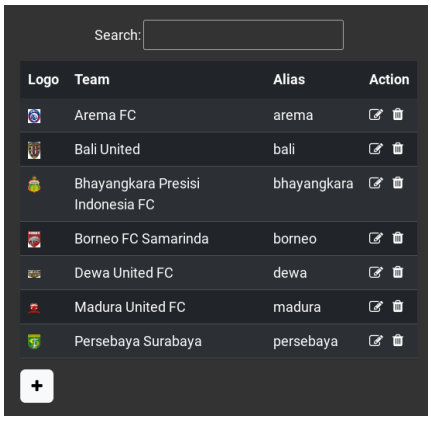
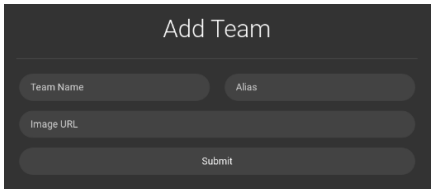
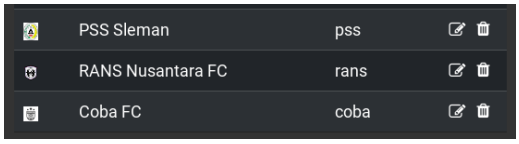
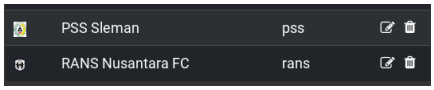
Figure 17. KNN Result

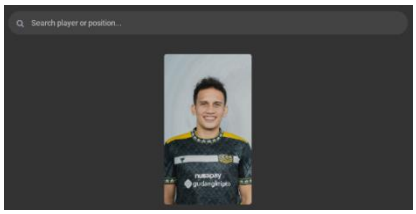
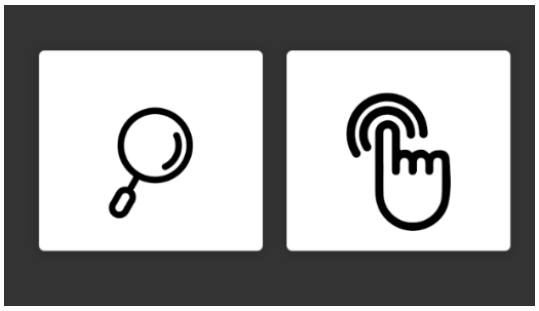
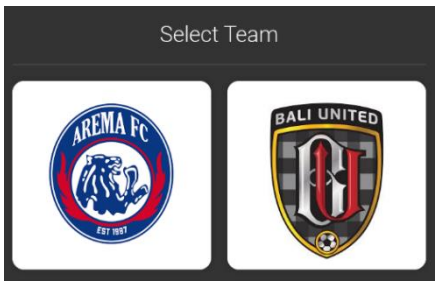
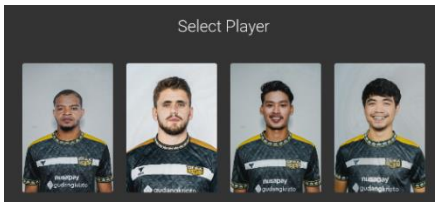
Based on the results above, the player named Egy Maulana Vikri received player recommendations, namely Renan Silva, Ricky Cawor, and Ariel Lucero.

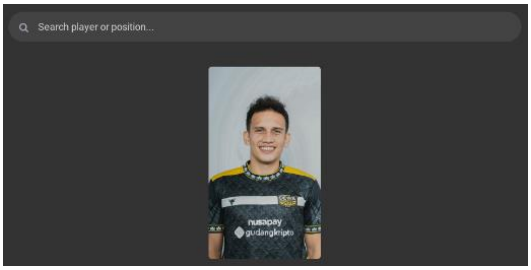
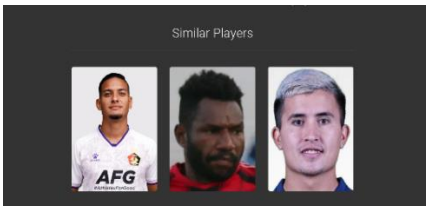
3.5. Testing

The data mining application program for player statistics data for scouting in this soccer game uses the blackbox method for testing. This method only performs testing based on the function of the program to ensure that it can run properly and correctly. The blackbox method testing is carried out by providing input data, processing it according to the function to see whether the program produces output according to the desired basic function. If the output is in accordance with its function, then the program can run properly and correctly, if not, there are still errors in the program and it needs to be fixed.

Tabel 2. *Testing*

Test Scenario	Expected Results	Results
The admin navigates to one of the database list pages, namely teams, players, and users (Example using teams). 	The database team list page opens according to the admin's choice. 	Valid
Admin opens the database list page according to the selected one (teams, players, or users). Example using teams.  Admin chooses to navigate the page to add, edit, or delete data (example using add)	The system displays the database related to the admin's choice.  The system displays the route according to the admin's choice, namely add. 	Valid
Admin clicks on the delete data option listed on the database page. 	The system processes the data and deletes the admin's selected data from the database and returns the page display to the database list. 	Valid

Test Scenario	Expected Results	Results
Users enter the search player page and type in the player of their choice or the desired player position. 	The system searches for data according to user selection and displays the players according to what the user has typed. 	Valid
The user enters the search player page and selects the relevant team. 	The system displays a team page containing all Indonesian League 1 teams competing in the 2023/2024 season.  The system displays the player page according to the selected team. 	Valid

Test Scenario	Expected Results	Results
Select the player of choice (select Egy Maulana Vikri). 	The system displays the statistical data of the selected player and gets three recommended players that match the player.   	Valid

4. CONCLUSION

Based on the results of the research, discussion and design that has been carried out related to the Application of Data Mining to Player Statistical Data for Scouting in Football Games at Dewa United, there are conclusions that are described as follows: 1. The scouting process carried out by the Dewa United FC team involves the scouting division going into the field to see talented players who want to be recruited and also get offers from registered agents. The results of the player's report are discussed again with the head coach for approval to the club president, 2. Obstacles that occur in the scouting process carried out by the Dewa United FC team are subjective observation results and are assisted by viewing raw data through the wyscout application. To solve these problems, the use of data that can provide scouting recommendations according to player statistics makes the scouting process effective and efficient, 3. Creating and implementing the use of player statistical data to provide a player recommendation system that matches the attributes selected by the user. The use of this data is made web-based which has key statistical application features and is comfortable for users to use.

5. SUGGESTED

Based on the results of the research that has been conducted, it is expected that this system can be further developed to be better, so that the scouting process can be more effective and efficient. The following are suggestions that can be considered: 1. In addition to going into the field, the scouting process can utilize an objective data-based recommendation

system so that the coach's chosen players can be right on target, 2. This research can be further developed by implementing anomaly detection and bid prediction and for the application of technology and greater convenience, it is necessary to implement a Large Language Model (LLM) which is limited to player recommendations so that player searches are much more targeted. The use of mobile devices is much greater than laptops or PCs today, making the use of Android or IOS-based data a consideration.

6. REFERENCES

[1] Cholisoh, Novi, Junaidi Junaidi, dan Irfa Safitri Sari. 2021. Rancangan Sistem Penginputan Judul Online KKP, TA/SKRIPSI Berbasis Laravel Pada Universitas Raharja. Technomedia Journal Vol. 5, No. 2:248-258.

[2] Asinjery.Wijoyo, Hadion. 2021. Sistem Informasi Manajemen. Solok: Insan Ce ndekia Mandiri.

[3] Maulani, Giandari, Handi Komara, dan Santi Meiliana. 2020. Rancang Bangun Sistem Informasi Monitoring Dashboard Traffic Work Order Berbasis Web. Journal Cerita Vol. 6, No. 2:137-146.

[4] Rahayu, Sri, Ilamsyah, dan Riena Eka Putri. 2019. Rancang Bangun Sistem Pengarsipan Data Jaminan Peserta (Studi Kasus BPJS Ketenagakerjaan Cikokol Tangerang). Jurnal Sistem Informasi dan Informatika (SIMIKA). Vol. 2, No. 2:1-12.

[5] Rahayu, Sri, Sandro Alfeno, dan Kartika Novianti Wahyono. 2019. Rancang Bangun Aplikasi Sistem Pengolahan Data Pembuatan Akta Tanah Pada Kantor Desa Cangkudu Tangerang. Innovative Creative and Information Technology. Vol. 5, No. 1:1-11.

[6] Rudi, dan Dedy Rahman Prehanto. 2020. Pengembangan Aplikasi Sistem Pengelolaan Data Prestasi Mahasiswa Berdasarkan Standar ISO/IEC 25010. Jurnal Manajemen Informatika. Vol. 11, No. 1.

[7] Wahono, Sri, dan Hapzi Ali. 2021. Peranan Data Warehouse, Software dan Brainware Terhadap Pengambilan Keputusan (Literature Review Executive Support System for Business). Jurnal Ekonomi Manajemen Sistem Informasi. Vol. 3, No. 2:225-239.

[8] Astutik, Ika Ratna Indra dan Mochamad Alfan Rosid. 2020. Buku Ajar Basis Data Untuk Informatika. Sidoarjo: Umsida Press.

[9] Rahayu, Sri, Putri Rachmanda, dan Adisa Lahmania Putri. 2022. Perancang an Sistem Informasi Pelayanan Pembuatan Surat Menyurat Berbasis Web Pada Kantor Desa Saga. Journal CERITA: ISSN 2461-1417.

[10] Feronika, Nanda. 2019. Karakteristik Data untuk Informasi yang Berkualitas. School of Information Systems - Binus University. Diambil dari: https://sis.binus.ac.id/2019/04/05/karakteristik-data-untuk-informasi-yang-berkualitas/. (8 Mei 2024).

[11] Amna, Wahyuddin S, I Gede Iwan Sudipa, Tri Andi E. Putra, Ahmad Jurnaidi Wahidin, Wara Alfa Syukrilla, Anindya Khrisna Wardhani, Nono Heryana, Tutuk Indriyani, dan Leo Wilyanto Santoso. 2023. DATA MINING. Jakarta: PT Global Eksekutif Teknologi.

[12] Jayanti, Asri, Asean, Baso Iping, Abdul Wahab, Abdurrahman Misno, Sri Wahyuni Hasibuan, Tasrim, Muhammad Sholahuddin, Saryanto, Siswandi Sululing, Ujang Syahrul Mubarrok. 2021. Metodologi Penelitian Bidang Muamalah, Ekonomi Dan Bisnis. Bandung: CV. Media Sains Indonesia.

[13] Duggal, Nikita. 2023. *What Is Data Processing: Cycle, Types, Methods,Steps and Examples*. Diambil dari: https://www.simplilearn.com/what-is-data-processing-article. (12 Mei 2024).

[14] Bhat, Adi. 2023. *Data Analysis in Research: Types & Methods*. Diambil dari: https://www.questionpro.com/blog/data-analysis-in-research/. (12 Mei 2024).

[15] Arora, Simran Kaur. 2023. *What is Data Analysis? Methods, Techniques& Tools*. Diambil dari: https://hackr.io/blog/what-is-data-analysis-methods-techniques-tools. (12 Mei 2024).

[16] Calzon, Bernardita. 2023. *Your Modern Business Guide To Data Analysis Methods And Techniques*. Diambil dari: https://www.datapine.com/blog/data-analysis-methods-and-techniques/. (12 Mei 2024).

[17] Ahram, Tareq, dan Waldemar Karwowski. 2022. *Human Factors in Robots, Drones and Unmanned Systems*. AHFE International Vol. 57.

[18] Khosrow-Pour, Mehdi. 2020. *Encyclopedia of Organizational Knowledge, Administration, and Technology*. Florida: IGI Global.

[19] Liou, Fuewen Frank. 2019. *RAPID PROTOTYPING AND ENGINEERING APPLICATIONS: A Toolbox for Prototype Development*. Florida: CRC PRESS.

[20] Zai, Charles. 2022. Implementasi Data Mining Sebagai Pengolahan Data. Jurnal Portal Data Vol. 2, No. 3.

[21] Pratama, Arya Satya, Suci Maela Sari, Maila Faiza Hj, Moh Badwi, dan Mochammad Isa Anshori. 2023. Pengaruh Artificial Intelligence, Big data dan otomatisasi terhadap kinerja SDM di Era digital. Jurnal Publikasi Ilmu Manajemen Vol. 2, No. 4:108-123.

[22] Eriana, Emi Sita, dan Afrizal Zein. 2023. Artificial Intelligence (AI). Purbalingga: CV. Eureka Media Aksara.

[23] Sidabutar, Hasudungan, dan Horasman Perdemunta Munthe. 2022. Artificial Intelligence dan Implikasinya Terhadap Tujuan Pembelajaran Pendidikan Agama Kristen. JMPK: Jurnal Manajemen Pendidikan Kristen Vol. 2, No. 2:76-90.

[24] Sidik, Ahmad Darmawan, dan Alek Ansawarman. 2022. Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning. *Formosa Journal of Multidisciplinary Research* Vol. 1 No. 3:559-568.

[25] Tjahjadi, Evan Valdis, Budy Santoso, dan Serwin. 2023. Klasifikasi Malware Menggunakan Teknik Machine Learning. Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer Vol. 2 No. 1:60-70.

[26] Amaratunga, Thimira. 2021. *Deep Learning on Windows*. New York: Springer.

[27] Verdhan, Vaibhav. 2021. *Computer Vision Using Deep Learning*. New York: Springer.

[28] Laskhmanan, Valliappa, Martin Gorner, dan Ryan Gillard. 2021. *Practical machine learning for computer vision.* Sebastopol: O'Reilly Media Inc.

[29] Sathyanarayanan, S., dan Sanjay Chitnis. 2022. A Survey of Machine Learning in Healthcare. Florida: CRC Press:1-22.

[30] Hayat, Abdul, Era Era Hia, dan Dyna Halimah Tusyadyah. 2019. Rancang Bangun Aplikasi Petty Cash Berbasis Web Menggunakan Framework Laravel Pada Pt Bekasi Asri Pemula Tbk. Innovative Creative and Information Technology Vol. 5, No. 2:130-140.

[31] Nugroho, Asep Hardiyanto, dan Toyib Rohimi. 2020. Perancangan Aplikasi Sistem Pengolahan DataPenduduk Dikelurahan Desa Kaduronyok Kecamatan Cisata, Kabupaten Pandeglang Berbasis Web. Jutis (Jurnal Teknik Informatika) Vol. 8, No. 1:1-15.

[32] Lestari, Merryana, dan Eko Sediyono. 2021. Analisis dan Perancangan Sistem Informasi Panduan Kesehatan Balita. *Journal of Information Systems and Informatics* Vol. 3 No.1:83-95.

[33] Irmayana, Andi, dan Komang Aryasa. 2019. Sistem Informasi Penentuan Jumlah Produksi Emas Murni Antam Dengan Menggunakan Metode Trend Linear Berbasis Android. SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi Informasi Vol. 8 No. 2.

[34] Rinaldy, Rio. 2022. Analisis Bisnis Internal Menggunakan Metode Critical Success Factors. Jurnal Ekonomi Teknologi dan Bisnis (JETBIS) Vol. 1, No. 3:146-151.