2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS)

Unsupervised Method to Analyze Playing Styles of EPL Teams using Ball Possession-position Data

Pranav Verma
National University of Ireland Galway
pranavverma1295@gmail.com

Bharath Sudharsan
Data Science Institute
National University of Ireland Galway
bharath.sudharsan@insight-centre.org

Bharathi Raja Chakravarthi
Data Science Institute
National University of Ireland Galway
bharathi.raja@insight-centre.org

Colm O'Riordan
National University of Ireland Galway
colm.oriordan@nuigalway.ie

Seamus Hill
National University of Ireland Galway
seamus.hill@nuigalway.ie

***Abstract*—In the English Premier League (EPL) matches, a network of advanced systems gathers sports data in real-time to build a possession-position dataset. In this work, data fields from the sophisticated raw possession-position dataset were extracted and processed to build a transformed version of the raw dataset. This transformed version contains ball possession data from 3 areas and 9 zones of the pitch. Two experiments were run based on this transformed dataset, aiming to understand and analyze the playing styles of EPL teams. The analysis answers multiple questions such as, is the playing style of the top 3 teams (Manchester City, Liverpool, and Chelsea) same in both home and away matches, do away match conditions affect the playing style of teams, etc. Existing studies use multiple parameters such as goal scoring patterns, player performances, team performance, etc. to understand and analyze the playing style of teams. In this work, using just the ball possession-position data, the playing styles of teams were able to be derived. This reduces the usage of such multiple parameters to perform the same task, which is to understand and analyze the playing styles of teams.**

***Index Terms*—k-means, ball possession-position data, EPL, sports analytics, data transformation, elbow method, similar playing style**

I. Introduction

From then to now every football team has its distinctive strategy which is a key for their victory. Such a unique strategy is ideally preserved within the team. Today's advanced player tracking systems [1] [2] [3] uses a network of wireless cameras and sensing hardware to capture live in-game data and player parameters such as speed, acceleration, distance traveled, etc. from hundreds of data points. Such an advanced tracking system gathers sports data in split-second intervals using which coaches obtain insights of other league team's playing styles [4]. Hence, it is not possible to preserve all aspects of a team's strategy within the team [5]. Ball pass accuracy, ball presences in attacking areas, additional passing alternatives, player's pass completion percentage, etc. can be determined from the ball possession data of a match. Hence, ball possession is one of the most important statistics to measure a team's performance in a game [6]. In this work, a raw possession-position dataset built from such advanced system is used. Experiments are performed using a transformed version of this raw dataset to provide deeper insights on, the playing styles of EPL teams, grouping teams based on their playing style, how away matches influence playing style of a team, etc.

The scripts written to extract data, process data and run experiments are provided in: https://github.com/pranavverma12/Analyze-playing-styles-of-EPL-teams-applying-k-means-on-ball-possession-position-data

Successful EPL teams utilize defenders to build-up play and maintain possession via short passes [7], permitting the attacking team to move the opponents around the pitch and wait for attacking opportunities to emerge. Also, the findings of [7] highlight, EPL teams mostly use their defenders for building the play which might lead to a goal. On the other hand, the findings of [8] [9] [10] convey that, every EPL team has a unique playing style. These article contradicts the findings of [7]. Since various research outcomes contradict each other, there is a strong need for an accurate understanding of a team's playing style to find out weather dominating ball possession during matches may or may not be a feature of successful EPL teams. Research investigations convey that the home teams tend to perform a higher number of attacking actions [5] and have more ball possession than away teams [11]. When teams are playing away from home, positive and passionate support from the audience triggers positive momentum and favors the home team [5]. Hence, the pitch location and crowd influence the playing style of the team [5]. But investigations also say invariable of the pitch location, higher-ranked teams have less variation in performance and playing style than lower-ranked teams [12]. Again, since various research outcomes contradict each other, more accurate analysis is required to determine whether pitch location influences the playing styles of EPL teams.

In this work, data fields from the sophisticated raw possession-position dataset are extracted using the technique discussed in Section. II-B, followed by data pre-processing as discussed in Section. II-C to build a transformed version of the raw dataset named ball possession-position dataset. This transformed ball possession-position dataset is used to perform two experiments. The first experiment from Section. IV uses data from 3 areas of the pitch (defense, midfield, and attack) to cluster home and away teams with similar playing styles. The second experiment from Section. V is conducted using the 9 zones data which was derived by partitioning the data from 3 areas of the pitch. This second experiment is performed using the same k-means clustering technique used in experiment 1. The aim of both the experiments is to understand and analyze the playing styles of teams. The discussions based on analysis brings multiple valuable insights which are listed along with the experiments in Section IV & V.

II. Dataset

A. Dataset structure

The raw possession-position dataset is built by capturing data using a state-of-the-art camera and sensor technology from all the English Premier League (EPL) games 2017-2018 contains data records of 210 EPL matches played from August 2017 to January 2018. This dataset contains separate folders for each round played, which sums up to 2x folders. Inside every folder, there are 1x sub-folders with game_id as their folder names. Finally, inside every game folder, there are 3 different files as listed. The dataset and its source cannot be exposed due to its proprietary nature.

(a) Data file: The .DAT file contains the complete data recorded from before the beginning of the match till the end. Data was captured and recorded in intervals of one second by 25 different cameras set up across the stadium. Before the match begins, this camera-based data collection system can track around 30 people on the pitch in real-time. Once the match starts, the capacity of real-time tracking will be reduced to 24 people (22 players and 2 referees). The main fields in the .dat file are frame number of the camera, player details, and ball details. These fields are separated by a colon (:).

(b) eXtensible Markup Language file: This XML file contains information about the football pitch in terms of length of the pitch, the height of the pitch, match date, match time, game_id, starting & ending frame number for the first half, starting & ending frame number for the second half and finally the tracking area covered by the cameras.

(c) Physical split CSV file: This file contains individual details of physical activities performed by players during the match. Total distance, standing distance, high-speed running distance, sprinting, jogging count, running count and name of the player's teams are the data fields.

B. Data extraction

Initially, the ball possession co-ordinates, its corresponding camera frame number, pitch details and playing team names are extracted from the raw possession-position dataset to build a transformed version of the raw dataset named ball possession-position dataset. Since these required data fields are inside multiple nested folders, it's required to iterate into multiple folders from the starting directory to fetch data. The raw dataset size is approximately 8 Gb and the .DAT Files of each match consist of 176,907 rows of data. Combining these rows with other XML and CSV's files multiples the numbers of iterations. Extracting data from such a wast number of rows for 210 matches is one of the challenging tasks faced in this work. The complex structure of the dataset demands high computation resources to perform the extraction in a single run. One Round (10 matches) of data is extracted at a single time and stored in a separate CSV file. To extract the data of our interest, a python script leveraging libraries such as Os, Numpy, Pandas, and concepts from DOM (Document Object Model) is written. When this python script is executed, it creates a new CSV file with the required ball possession-position data for one round of matches. To build the complete transformed version of the raw dataset, this script has to be incrementally iterated 21 times to create separate CSV files for each round of matches. Each CSV file contains approximately 500 Mb of ball possession-position data.

C. Data pre-processing

The data fields for the new ball possession-position dataset (transformed raw possession-position dataset) is extracted using the Python script described in Section. II-B. At this stage, the ball possession-position dataset contains raw data fields. These raw data fields have to be processed to get converted into fields that can be used for the experiments. A single base CSV file from 21 separate CSV files is created. This base file contains transformed data fields from all the 210 matches played in 21 rounds. These transformed data fields are, match_id, ball co-ordinates of team A, ball co-ordinates of team B, names of team A (home teams), names of team B (away teams), zone data for team A, zone data for team B. The zone data columns are newly created. These columns contain ball possession percentage of the playing teams (team A & team B). To calculate the zone data for these new columns, initially, the pitch where the match was played is plotted, followed by computing the ball possession count in each zone of the pitch and transforming these counts into ball possessions percentage for each zone. At this stage, the data fields of the ball possession-position dataset are completely transformed and ready to be used for experiments.

III. Methodology

The k-means algorithm is one of the most used unsupervised learning algorithms to solve the well-known clustering

Fig. 1. 9 zones and 3 areas of a pitch

problem. A cluster is collection of data objects that are homogeneous within one cluster and heterogeneous to object in other cluster. If we want to group input vectors into K clusters, K-means can output clusters of $w_1,....w_k$ and their corresponding vectors $\mu_1....\mu_k$ such that they locally minimize the residual sum of squares (RSS) which is defined as

$$J = \sum_{k=1}^{K} \sum_{x \in w_k} ||x - \mu_k||^2 \quad (1)$$

In the algorithm, μ_k is made into the mean of the vectors in a cluster w_k. Hence comes the name K-means. This means that the k-means algorithm tries to optimize the objective function 1. As there is only a finite number of possible assignments for the amount of centroids and observations available and each iteration has to result in better solution, the algorithm always ends in a local minimum [13].

In this work, the k-means algorithm is applied to the data fields from the ball possession-position dataset to cluster teams with a similar playing style. k-means is chosen over other algorithms since it is simple to interpret the clustering results, fast and efficient in terms of computational cost and generalizes to clusters of different shapes and sizes. In K-means, a cluster refers to a collection of data points aggregated together because of certain similarities [14]. In this case, a cluster contains a group of teams with a similar playing style. A centroid is an imaginary or real location representing a center for each of the formed clusters [14]. In this case, a centroid is a center point of a cluster containing grouped teams with similar playing styles. The k-means clustering algorithm used in both the experiments identifies k number of centroids, and then allocate every team to the nearest cluster of teams while keeping the centroids as near as possible. An elbow curve method is used in both the experiments to assist the selection of the optimal number of clusters by fitting the model with a range of values for k. Fig. 2 & Fig. 6 are outputs of k-means clustering using the elbow method. Here, from the curves, a point of inflection represents an "elbow" of an arm. The optimal number of clusters is the values corresponding to these elbow points [15]. This Elbow method is implemented using WCSS since it is a technically robust method [16]. In upcoming Sections, two

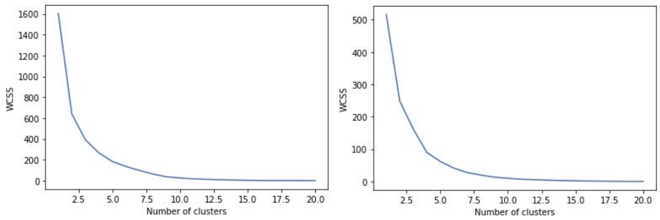

Fig. 2. Elbow curve to find optimal number of clusters for experiment 1

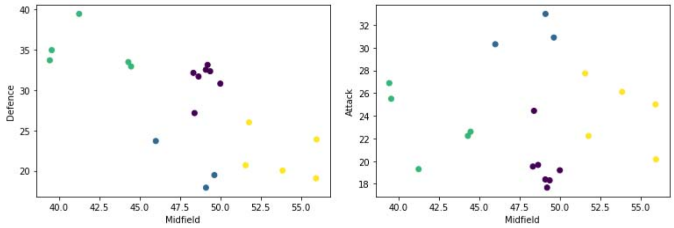

Fig. 3. Clustered home teams with a similar playing style using data from 3 areas of the pitch

experiments are performed based on the ball possession-position dataset, using k-mans algorithm. The aim of the experiments are to understand and analyze playing styles of teams. The discussions based on analysis brings multiple valuable insights which are listed in upcoming sections.

IV. Results and Discussion: Experiment 1

As shown in Fig. 1, the pitch contains 3 areas which are further divided into 9 zones. Zones 1-3 forms the defense area; zones 4-6 form the midfield area and zones 7-9 form the attacking area. The first subpart of this experiment uses data from 3 areas of the pitch (defense, midfield, and attack) to cluster home teams with similar playing styles. The second sub-part of this experiment uses the same 3 area data to cluster away teams with a similar playing style. In Fig. 2, the x-axis ranges from 1 to 20 because a minimum number of the cluster can be one and the maximum number of clusters can be twenty. Values of Within-Cluster-Sum-of-Squares (WCSS) are on the Y-axis which defines the reduction in the cost during training. Fig. 2 (left) shows that the curve is bending at point 4 (elbow point) for home teams. Similarly, the elbow point is 4 from Fig. 2 (right) for away teams. These elbow points are selected since it achieves reasonable performance without having many or less number of clusters [16].

A. Cluster home teams with similar playing style using data from 3 areas of the pitch

The 4 clusters of home teams with a similar playing style are shown in Fig. 3 and in Table. I. The discussion based on obtained results are listed

(a) The top 10 league teams of the EPL season 2017-18 is provided at [17]. The team names from cluster 3 are the same as the top 7 teams mentioned in [17]. The match of obtained clustering results with the known fact strongly supports the exactness of experiment results.

(b) From Fig. 3, it can be observed that the teams from cluster 3 are more scattered and teams from this cluster have more ball possession percentage in the midfield area. Also, the teams from cluster 3 have maximum ball possessions percentage in the attack area.

(c) The teams from cluster 2 are most tightly grouped, hence teams from this cluster do not have much variation in their playing style.

(d) Cluster 1 and cluster 0 contains outliers since teams are not closely grouped. The reason is that cluster 1 contains teams that are at the 7th position (Wolverhampton Wanderers), 13th position (Newcastle United) and 16th position (Southampton). Similarly, Cluster 0 contains outliers since Tottenham Hotspur and Arsenal are from the top 5 places of the ranking table, whereas Burnley and Fulham are from the bottom of the ranking table.

(e) According to [9] Arsenal and Burnley are building their attacks from the defense area. According to Fig. 3, Arsenal and Burnley have high ball possession percentage in the defense area since these teams are from cluster 0. Again, the match of obtained clustering results with the known fact strongly supports the exactness of experiment results.

TABLE I
The 4 cluster's of home teams with a similar playing style obtained using data from 3 areas of the pitch

Cluster	Colour	Team names
0	Green	Tottenham Hotspur, Arsenal, Leicester City, West Ham United, Watford, Burnley, Fulham
1	Blue	Wolverhampton Wanderers, Newcastle United, Southampton
2	Violet	Crystal Palace, Bournemouth, Brighton Hove Albion, Cardiff City, Huddersfield Town
3	Yellow	Manchester City, Liverpool, Chelsea, Manchester United, Everton

B. *Cluster away teams with a similar playing style using data from 3 areas of the pitch*

The four cluster's of away teams with a similar playing style are shown in Fig. 4 and Table. II. The discussion based on obtained results are listed

(a) From Fig. 4, it can be observed that the teams from cluster 3 are scattered. It is because this cluster has teams with mixed rankings. Everton is ranked 7th, Newcastle United from the middle of the ranking table and Cardiff City at the bottom of the ranking table. In away games, these three teams have maximum ball possession in the defense area as compared to other cluster teams and their counter-play starts from the defense area. While in home matches their counter-play was mostly started from midfield. Likewise, in away matches, Tottenham Hotspur and Southampton have more ball possession percentages in attacking areas. Whereas in home matches, they have more ball possession percentages in the mid-field area. This brings to the conclusion that Everton, Newcastle United, Cardiff City, Tottenham Hotspur, and Southampton have different playing styles in away matches when compared with home matches.

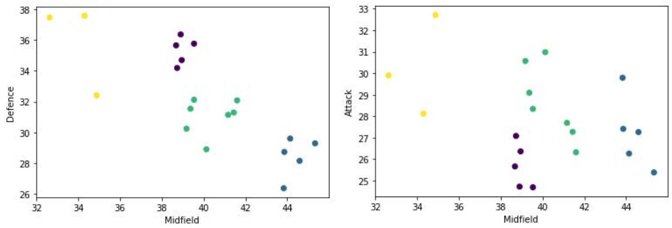

Fig. 4. Clustered away teams with a similar playing style using data from 3 areas of the pitch

(b) According to [17], there is not much variation in the game-play of the top 3 teams (Manchester City, Liverpool, and Chelsea) of EPL season 2017-2018. From Fig. 3 & Fig. 4 it can be observed that for the top three teams, the majority of ball possession is in attacking areas of the pitch for both home and away matches. Again, the match of obtained clustering results with the known fact strongly supports the exactness of experiment results

(c) In Cluster 1, except for Southampton, other teams are in the top 4 places of the ranking table. Whereas Southampton is from the bottom of the table. This result can bring to an inference that even low ranked teams can have playing style similar to top-ranked teams

On the basis of discussions from Section. IV-A and Section. IV-B, it is concluded that

(a) The playing styles of teams (Everton, Newcastle United, Cardiff City, Tottenham Hotspur, and Southampton) change in the away matches when compared with home matches.

(b) The playing style of top 3 teams (Manchester City, Liverpool, and Chelsea) are the same in both home and away matches as they have high ball possession percentage in the attack area.

(c) Ball possession data from 3 areas of pitch and game location (home or away) data is sufficient enough to cluster similar teams.

V. Results and Discussion: Experiment 2

Clustering of home and away teams using data from 3 areas of the pitch was performed in Experiment 1 and

TABLE II
The 4 cluster's of away teams with a similar playing style obtained using data from 3 areas of the pitch

Cluster	Colour	Team names
0	Violet	Wolverhampton Wanderers, Crystal Palace, Burnley, Fulham, Huddersfield Town
1	Blue	Manchester City, Liverpool, Chelsea, Tottenham Hotspur, Southampton
2	Green	Arsenal, Manchester United, Leicester City, West Ham United, Watford, Bournemouth, Brighton and Hove Albion
3	Yellow	Everton, Newcastle United, Cardiff City

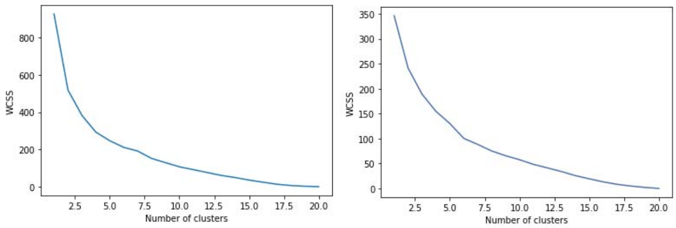

Fig. 5. Elbow curve to find optimal number of clusters for Experiment 2

basic insights were obtained based on the results. Deeper analytics to obtain additional insights on playing styles of teams, weak & strong areas of the team, etc could not be obtained from the results of Experiment 1. Hence, there is a need to conduct another experiment using the 9 zones data which was derived by partitioning the data from 3 areas of the pitch. This second experiment is performed using the same k-means clustering technique used in experiment 1. The first sub-part of this experiment uses data from 9 zones of the pitch to cluster home teams with a similar playing style. The second subpart performs the same clustering, using the same 9 zone data but for away teams. The elbow curves to find the optimal number of clusters were plotted using data from 9 zones of the pitch. From Fig. 5, the elbow point for both the home and away team was found to be 4.

A. Cluster home teams with a similar playing style using data from 9 zones of the pitch

For the host/home team, the pitch is divided into zones starting from the left bottom corner as shown in Fig. 1. Clustering was performed and the result containing 4 clusters of home teams with similar playing style are shown in Table. III and a part of the resulting scatter plot is shown in Fig. 6. From this scatter plot it can be observed that only 17 teams out of 20 teams are visible. Huddersfield Town is

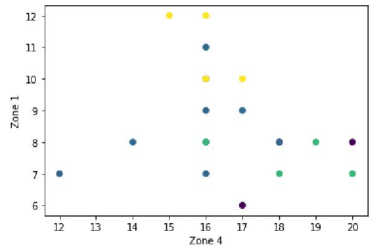

Fig. 6. Clustered home teams with a similar playing style using data from 9 zones of the pitch

overlapping with Tottenham Hotspur, followed by Chelsea with Westham United and Watford with Newcastle United. Due to this overlap, the discussion cannot be made based on the scattered plot. Rather the discussion is made based on the Table. IV.

TABLE III
The 4 cluster's of home teams with similar playing style obtained using 9 zones data of the pitch

Cluster	Colour	Team names
0	Violet	Manchester City, Liverpool, Chelsea
1	Blue	Tottenham Hotspur, Arsenal, Manchester United, Leicester City, West Ham United, Watford, Crystal Palace, Burnley, Fulham
2	Green	Wolverhampton Wanderers, Everton, Newcastle United, Southampton
3	Yellow	Bournemouth, Brighton and Hove Albion, Cardiff City, Huddersfield Town

B. Cluster away teams with a similar playing style using data from 9 zones of the pitch

For the guest/away team, the pitch division starts from the right bottom corner. The four clusters of away teams with similar playing styles are shown in Table. V. A scatter plot is not used due to the reason discussed in Section. V-A. Rather the discussion is made based on the Table. VI

C. Discussion

Based on clustering results and tabulations from Section. V-A and Section. V-B, the listed insights are obtained

(a) All the teams from Cluster 3 of Table. III is from the bottom of the official ranking table [17]. Bournemouth (13th place), Brighton & Hove Albion (16th place), Cardiff City (17th place) and Huddersfield at the last place. From Table. IV it can be derived that these low ranked teams have more ball possession at the central back (also known as a central defender or center-half) than at the full-backs (the left-back and the right-back).

TABLE IV
Home teams: Ball possession percentage in each of the 9 zones of the pitch

Teams	z1	z2	z3	z4	z5	z6	z7	z8	z9
Manchester City	6	8	5	17	19	21	8	7	10
Liverpool	8	9	7	20	16	20	7	7	6
Chelsae	8	5	7	18	19	18	7	10	8
Tottenham Hotspur	10	12	9	16	15	18	6	6	7
Arsenal	11	13	9	16	14	19	6	5	7
Manchester United	7	11	8	16	16	20	6	7	10
Wolverhampton Wanderers	7	5	6	18	14	17	8	16	8
Everton	7	5	8	20	14	18	9	11	8
Leicester City	9	13	9	16	14	18	6	7	7
West Ham United	8	12	7	18	14	16	8	9	8
Watford	8	12	12	16	15	18	6	6	6
Crystal Palace	7	15	11	12	11	21	7	6	9
Newcastle United	8	7	8	16	14	16	8	13	9
Bournemouth	10	13	10	17	13	15	8	8	6
Burnley	8	15	8	14	15	21	7	7	5
Southampton	8	6	6	19	14	17	8	12	10
Brighton and Hove Albion	12	17	11	15	11	15	7	6	6
Cardiff City	12	15	8	16	12	11	8	9	9
Fulham	9	15	8	17	16	16	6	7	6
Huddersfield Town	10	14	9	16	10	13	10	10	7

TABLE VI
Away teams: Ball possession percentage in each of the 9 zones of the pitch

Teams	z1	z2	z3	z4	z5	z6	z7	z8	z9
Manchester City	8	11	8	15	14	15	8	12	9
Liverpool	8	13	9	15	13	16	8	12	7
Chelsea	7	12	8	14	16	14	9	12	7
Tottenham Hotspur	9	12	9	15	15	15	6	11	8
Arsenal	9	15	8	14	14	14	7	12	7
Manchester United	7	15	9	13	15	14	7	12	8
Wolverhampton Wanderers	8	18	10	13	13	13	6	12	7
Everton	9	14	9	12	11	12	8	16	9
Leicester City	7	13	9	13	12	15	9	14	8
West Ham United	7	14	9	13	12	14	6	16	8
Watford	9	15	8	14	13	13	8	13	8
Crystal Palace	8	18	9	14	13	11	7	12	7
Newcastle United	9	17	11	12	10	11	10	12	6
Bournemouth	8	15	9	14	14	13	7	13	8
Burnley	10	16	9	13	12	14	6	14	7
Southampton	8	15	7	16	14	14	8	12	7
Brighton and Hove Albion	9	13	9	12	10	18	8	12	8
Cardiff City	10	16	11	12	9	12	8	13	8
Fulham	9	19	8	12	12	15	8	12	6
Huddersfield Town	8	17	10	12	13	15	6	11	8

TABLE V
The 4 cluster's of away teams with similar playing style obtained using data from 9 zones of the pitch

Cluster	Colour	Team names
0	Violet	Arsenal, Manchester United, Everton, Leicester City, West Ham United, Watford, Bournemouth, Burnley, Southampton
1	Blue	Newcastle United, Cardiff City
2	Green	Manchester City, Liverpool, Chelsea, Tottenham Hotspur, Brighton and Hove Albion
3	Yellow	Wolverhampton Wanderers, Crystal Palace, Fulham, Huddersfield Town

(b) Cluster 2 from Table. III contains 2 teams from the top 7 ranks (Everton & 'Wolverhampton Wanderers) and 2 teams (Newcastle United & Southampton) from the middle of the official league ranking table [17]. For all these four teams, ball possession percentage in their mid-wings (right and left wings of the mid area) is more than that at the central midfield.
From Table. IV it is clear that Crystal Palace from Cluster 1 of Table. III uses its left-midfield wing area more when compare to other midfield areas. Likewise, other teams in this Cluster such as Leicester City and Burnley also have higher ball possession in the left-midfield wing area.

(c) Cluster 0 from Table. III represents the top 3 teams of the league (Manchester City, Liverpool, and Chelsea). From Table. IV, Man City and Liverpool use the wing areas of the pitch to build and strengthen their attacks. In contrast, Chelsea uses the center of the pitch to build and strengthen their attacks.

(d) From Table. VI, it can be observed that tactics of the top 4 teams in Cluster 2 of Table. IV has been changed in away matches. The central-attacking zone (Zone 8) of the pitch has more ball possession in away matches. Whereas in home matches, the wings of the attack area have more ball possession.

(e) Cluster 1 from Table. IV contains 2 teams (Newcastle United and Cardiff City). Table. VI shows, the ball possession of these two teams in defense zones (Zone 1 to Zone 3) is almost similar.

When comparing the discussions from Section. IV-A and Section. IV-B of Experiment 1 with the discussion from Section. V-C of Experiment 2. All the points from discussions based on Experiment 2 provide information conveying which particular zones from the three areas of the pitch are more used by teams, rather than saying teams had more ball possession in the defense area. It can be concluded that using data from 9 zones to perform clustering provides more insightful discussions in terms of playing styles of teams, weak & strong areas of the team.

VI. Conclusion

Two experiments were run based on the transformed dataset and discussions based on the experimental results provided a solid understanding of playing styles of teams. The outcomes of the discussions bring multiple valuable insights. Firstly, the playing style of top 3 teams (Manchester City, Liverpool, and Chelsea) are the same in both home and away matches and they have high ball possession percentage in the attack area than in other areas. Next, the ball possession data from 3 areas of pitch and game location (home or away) data is sufficient enough to cluster similar teams. Next, when clustering is performed using data from 9 zones of the pitch, more insightful discussions in terms of playing styles, weak & strong areas of the teams are obtained. Finally, the analysis favors the known fact that away match conditions affect the playing style of a few teams as their tactics differ from home and away matches. To conclude, using just the ball possession position data, the playing styles of teams were able to be derived. This reduces the usage of multiple parameters such as goal scoring patterns, player performances, team performance, etc. to perform the same task, which is to understand and analyse playing styles of teams.

VII. future directions

Few more data fields from the raw dataset are planned to be used. Future work concerns to use these new data fields as input for the same k-means clustering algorithm to plot the scatter chart from which deeper analysis could be performed to understand how ball possession influences the match results. In the current study, the pitch was divided into 9 zones of equal size. In the future, the zones from left-wing can be grouped covering 40% of the pitch. Likewise, the zones from right-wing can be grouped covering another 40% of the pitch. The remaining 20 % from the zones in the central part of the pitch. The same experiments could be run based on the ball possession from the new pitch divisions.

References

[1] "Optical player tracking | player tracking system | stats," 2019. [Online]. Available: https://www.stats.com/player-tracking/

[2] "Electronic performance tracking systems," 2015. [Online]. Available: https://football-technology.fifa.com/en/media-tiles/epts-1/

[3] L. Tedesco, "Sports tracking – chyronhego," 2018. [Online]. Available: https://chyronhego.com/products/sports-tracking/

[4] H. Sarmento, A. Pereira, N. Matos, J. Campaniço, T. M. Anguera, and J. Leitão, "English premier league, spaińs la liga and italýs seriés a – what's different?" *International Journal of Performance Analysis in Sport*, vol. 13, pp. 773–789, 12 2013.

[5] A. J. Redwood-Brown, P. G. O'Donoghue, A. M. Nevill, C. Saward, and C. Sunderland, "Effects of playing position, pitch location, opposition ability and team ability on the technical performance of elite soccer players in different score line states," *PLOS ONE*, vol. 14, p. e0211707, 02 2019.

[6] "Keep the ball! the value of ball possession in soccer," 2010. [Online]. Available: https://warwick.ac.uk/fac/cross_fac/iatl/reinvention/archive/volume6issue1/parzialeandyates

[7] D. Adams, R. Morgans, J. Sacramento, S. Morgan, and M. D. Williams, "Successful short passing frequency of defenders differentiates between top and bottom four english premier league teams," *International Journal of Performance Analysis in Sport*, vol. 13, pp. 653–668, 12 2013.

[8] L. Gyarmati, H. Kwak, and P. Rodriguez, "Searching for a unique style in soccer," 2014. [Online]. Available: https://arxiv.org/abs/1409.0308

[9] "Espn," 10 2018. [Online]. Available: https://www.espn.com/soccer/english-premier-league/23/blog/post/3647791/premier-league-style-of-play-rankings-put-manchester-united-in-the-danger-zone

[10] "r/soccer - can anyone describe each epl team's playing style?" 2011. [Online]. Available: https://www.reddit.com/r/soccer/comments/rqaul/can_anyone_describe_each_epl_teams_playing_style/

[11] C. Lago and R. Martín, "Determinants of possession of the ball in soccer," *Journal of Sports Sciences*, vol. 25, pp. 969–974, 07 2007.

[12] C. Lago-Peñas and A. Dellal, "Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables," *Journal of Human Kinetics*, vol. 25, pp. 93–100, 09 2010.

[13] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, 1967, pp. 281–297. [Online]. Available: https://projecteuclid.org/euclid.bsmsp/1200512992

[14] M. J, "Understanding k-means clustering in machine learning," 09 2018. [Online]. Available: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

[15] "Elbow method — yellowbrick 0.9.1 documentation," 2016. [Online]. Available: https://www.scikit-yb.org/en/latest/api/cluster/elbow.html

[16] "Edupristine," 11 2018. [Online]. Available: https://www.edupristine.com/blog/beyond-k-means

[17] NDTVSports.com, "English premier league 2018-19 - table and standings | ndtvsports.com," 2018. [Online]. Available: https://sports.ndtv.com/english-premier-league/epl-table/2018-19