open access

Published online on: https://journal.uii.ac.id/jurnalsnati/

Check for updates

DOI: 10.20885/snati.v4.i2.40910

JURNAL SNATI Sains, Nalar, dan Aplikasi Teknologi Informasi

Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi

Vol. 4 No. 2 (2025) 54 - 61 ISSN Media Electronic: 2807-5935

Mapping the Safest Routes: A Clustering Study of the French Defense

Feri Wijayanto
Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia
feri.wijayanto@uii.ac.id

Abstract

This study explores the safest variations in the French Defense using 5,156 artificially generated chess games with Stockfish 17. Unlike prior work reliant on historical data, this method reduces theory bias by randomly selecting from the engine's top five moves at each position. We applied k-means clustering with cosine similarity to group move sequences based on evaluation scores. Both two-cluster and three-cluster models were tested. Stability was assessed via 50 resamples using 50% of the data. The three-cluster model, which includes a neutral group, had excellent stability (ARI = 0.99) but moderate cohesion (silhouette = 0.53). The two-cluster model showed better cohesion (silhouette = 0.65) but lower stability (ARI = 0.68). Among the variations, e5 (Advance) and exd5 (Exchange) stood out, with about 54% of games in each line falling into clusters favoring White. This suggests they are the safest and most reliable options. In contrast, Bb5+ performed well in simulations but poorly in real-world data, indicating theoretical risks. In summary, clustering on simulated games reveals hidden strategic insights, confirming e5 and exd5 as strong, low-risk choices for White in the French Defense.

Keywords: *clustering analysis; k-means with cosine similarity; cluster stability with resampling; adjusted rand index; silhouette index*

1. Introduction

Chess is not only a competitive game but also a complex system of decision-making that has attracted extensive research across domains such as artificial intelligence, strategy modeling, and pattern recognition. One of the most studied phases in chess is the opening, which sets the foundation for the middle and endgame. Understanding the dynamics of chess openings is crucial because the initial moves often determine the trajectory of the game, influencing control of the center, development, king safety, and tempo [1].

This study focuses on the French Defense, initiated by the move sequence *1.e4 e6 2.d4 d5* [2]. The choice of French Defense in this research was arbitrary, yet it remains one of the most frequently played responses to 1.e4 and has a rich theoretical foundation. The popularity of the French Defense makes it a suitable subject for experimentation while still offering sufficient diversity in variations to explore strategic tendencies.

A prevailing belief in chess theory is that White inherently possesses a slight advantage due to moving first [3]. However, this advantage is not unconditional: if White makes suboptimal moves in the opening, the initiative can shift rapidly in favor of Black. The sequence of these alternating moves forms a branching structure, where each decision influences the next. These interconnected sequences are known as variations. A variation in chess refers to a specific sequence of moves—chosen by both players—that follows a particular path from the starting position. Because of the vast number of legal moves and possible replies, the number of potential variations is enormous. The study of these variations, especially in the early phase of the game, is known as opening theory. With numerous theoretical and practical lines available, a natural question arises: which variations can be considered the safest or most favorable for White?

Previous studies have primarily relied on real-world chess game datasets, such as grandmaster games or online databases like chess.com or lichess.org, along the line [4], [5], [6], [7], [8]. While useful, such datasets are heavily influenced by established theory, opening preparation, and widely known strategic patterns. This introduces a knowledge bias and may obscure alternative strategies that lie outside the scope of common practice.

To address this gap, this study proposes a novel approach: generating artificial chess games using the Stockfish engine version 17. At each decision point, one of the top five recommended moves is selected at random, producing diverse yet high-quality game sequences. By escaping the boundaries of conventional theory, this approach aims to explore the full potential of White's early strategic choices and identify which

Received: 08-06-2025 | Revised: 01-07-2025 | Accepted: 02-07-2025 | Published: 10-07-2025

move sequences consistently yield advantageous positions.

To analyze the resulting dataset of generated game sequences, we employ a clustering technique. Specifically, k-means clustering is used in conjunction with cosine similarity, as in [9], [10], to group games according to the directional trend of positional evaluation scores. This combination allows the identification of latent patterns and structural similarities in move sequences that favor White or Black.

Several prior works have applied using chess games database using machine learning method as in [11], [12]. Moreover, on clustering to chess data, such as evaluation on Forsyth-Edwards Notation (FEN) structures using depth-based clustering [7], and investigation of Portable Game Notation (PGN) clustering [8]. However, these studies were limited to real-world data or specific representations, lacking the exploratory potential enabled by simulation.

In summary, this study aims to investigate which opening variations within the French Defense offer the safest routes for White by leveraging artificially generated chess games and unsupervised learning techniques. The findings are expected to shed new light on move safety and provide insights that are less constrained by traditional chess theory.

2. Research Methods

This study utilizes a dataset comprising 5156 artificially generated chess moves, each uniquely designed to explore an extensive range of potential moves. This methodological choice significantly departs from prior research, such as those described in previous studies [7], [8], which predominantly analyzed real-world chess games. Employing artificially generated moves allows the examination of a broader spectrum of possible chess scenarios that may not typically appear in conventional games.

The rationale for using artificially generated moves is to investigate possibilities beyond the constraints of established theories and commonly executed moves. Traditional studies typically focus on competitive matches and documented chess theories, thus overlooking unconventional or rarely explored strategies. By intentionally generating moves outside these standard boundaries, the research aims to identify patterns, strategies, or tactical nuances otherwise neglected, enriching the understanding of chess dynamics from a novel perspective.

Artificially generated moves further provide a controlled environment, minimizing biases inherent in historical data, such as repetitive patterns or popularized strategies. This unbiased dataset facilitates an objective evaluation of move characteristics, systematically addressing the complexity, safety, and strategic diversity inherent in chess. These introductory insights will subsequently be elaborated upon in dedicated sub-sections of the research methodology, including detailed explanations of data generation and preparation, distance computation methods, clustering algorithms, and evaluation metrics used in this study. The research methodology used in this study is summarized in Figure 1.

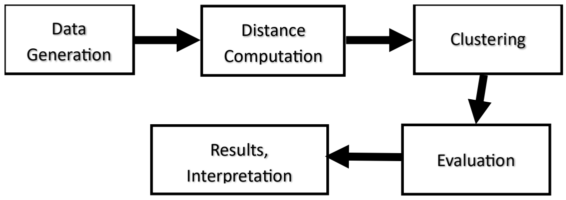

Figure 1. Research Methodology

2.1. Data Generation

The data generation phase of this research specifically targets the French Defense chess opening, beginning from the sequence *1. e4 e6 2. d4 d5*. To ensure comprehensive coverage of potential move sequences, the Stockfish chess engine (version 17) was employed. At each decision point during move generation, one of the five highest-ranking moves recommended by Stockfish was randomly selected, introducing variability and breadth into the generated data. Limiting the selection to the top five moves assumes professional players typically strive for optimal moves, and restricting choices to these highest-ranked options sufficiently captures the decision-making quality expected at advanced levels.

Each generated chess sequence was expanded to include a total of 30 additional moves beyond the initial opening sequence. This structure resulted in chess games consisting of 34 moves in total—17 moves for white and 17 moves for black—enabling thorough analysis of mid-game scenarios arising specifically from the French Defense.

In addition to the generated move sequences, positional evaluations provided by the Stockfish engine were recorded at each step. These evaluation scores form the analytical backbone of this study, serving as a fundamental measure for investigating strategic effectiveness, positional safety, and complexity within each generated scenario.

Figure 2 presents a sample of generated chess moves along with their evaluation scores from the Stockfish engine. The top row displays the sequence of moves, while the row beneath it shows the engine's evaluation for each position. When written in PGN format, the move sequence is: *1.e4 e6 2.d4 d5 3.e5 c5 4.Nd2 cxd4 5.Nb3 Bd7 6.Bd3 Ne7 7.f4 Nbc6.* Each move leads to a new board position; for example, after Black's move *7...Nbc6*, the resulting position is illustrated in Figure 3. To assess the quality of each position, Stockfish uses a metric called centipawn loss (CPL), which measures how much weaker a move is compared to the best move suggested by the engine. One centipawn equals one-hundredth of a pawn, so a CPL of 20 (as after *7...Nbc6*) means the move is one-fifth of a pawn worse than the optimal choice. Positive evaluation values suggest an advantage for White, while negative values favor Black. It is worth noting that CPL scores can differ between engines, as each engine uses its own algorithm and evaluation method to calculate the best moves and losses. As a result, CPL is dependent on the specific design of the engine being used. This evaluation system helps to highlight which positions are stronger or weaker throughout a series of moves in the game.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
1	e5	c5	Nd2	cxd4	Nb3	Bd7	Bd3	Ne7	f4	Nbc6
2	38	28	-22	4	-1	17	-1	22	18	20

Figure 2. Example of the generated chess move. First row indicates the move sequence and the second row indicates the correspondence position evaluation score.

Figure 3. The position resulting from the sequence of moves 1. e4 e6, 2. d4 d5, 3. e5 c5, 4. Nd2 cxd4, 5. Nb3 Bd7, 6. Bd3 Ne7, 7. f4 Nbc6.

Five distinct variations emerged through this randomized generation approach:

- 3. e5 (advanced variation),
- 3. exd5 (exchange variation),
- 3. Nc3 (Paulsen variation),
- 3. Nd2 (Tarrasch variation),
- 3. Bb5+ — Stockfish's fifth-best recommendation, which is absent from traditional French Defense theory.

The final dataset comprises a diverse set of scenarios, specifically 1126 variations beginning with e5, 1030 variations with exd5, 990 variations using Nc3, 1027 variations starting with Nd2, and 983 variations employing Bb5+. This broad, balanced selection of scenarios ensures robust analysis across conventional and unconventional strategic paths.

2.2. Distance Computation

In this research, the "safest route"—particularly from the perspective of the white side—is defined as the variant of the French Defense opening sequence that maximizes advantages for white. Specifically, a safer route for white is characterized by a higher frequency of favorable moves, as indicated by positive evaluation scores, coupled with fewer opportunities for advantageous moves by black, marked by negative evaluation scores.

The fundamental concept involves grouping sequences of moves based on the similarity of their evaluation score trends. By clustering these sequences, it becomes possible to identify patterns and variants that consistently favor white. Such clustering enables a clearer understanding of optimal strategic pathways within the complexity of chess openings, particularly highlighting sequences that systematically yield positional advantages to white.

From a technical perspective, accurately capturing these patterns necessitates selecting an appropriate measure of distance. Given that the dataset comprises positional evaluation scores with both positive and negative values, it is crucial not only to measure the magnitude of differences but also to effectively capture the directional aspect—positive versus negative scores. Consequently, cosine similarity has been employed as the distance measure for clustering [13], [14]. Cosine similarity effectively captures both magnitude and directional components, making it particularly suitable for this analysis, thereby ensuring the method effectively group routes exhibiting similar evaluation trends.

The cosine similarity formulized as

$$similarity(A, B) = \sin\theta = \frac{A \cdot B}{\|A\|\|B\|}$$

and the distance is formulized as

$$distance(A, B) = 1 - similarity(A, B),$$

where A is the sequence of the evaluation scores of the first game and B is the sequence of the evaluation scores of the second game.

2.3. Clustering Algorithm

This study employs k-means clustering, a straightforward and efficient clustering technique, to analyze the generated chess move sequences. Although hierarchical clustering is typically suited for identifying patterns of similarity among sequences, its focus primarily on physical distance metrics makes it less suitable for this analysis. Given that the directional aspect of evaluation scores—positive versus negative—is critical, the study utilizes cosine similarity to effectively capture both magnitude and direction of the evaluation trends. Consequently, k-means clustering, which accommodates cosine similarity more effectively, becomes the preferable method.

Initially, the analysis anticipates dividing the data into three distinct clusters: those favoring white, those favoring black, and an unclear or neutral cluster. This initial choice of three clusters (k = 3) aligns logically with the goal of distinctly categorizing game scenarios based on evaluation score trends. However, to ensure optimal clustering and robustness of analysis, the silhouette index will also be computed to validate or refine the choice of cluster numbers[15]. The silhouette index assesses how distinct data points have been grouped, providing a data-driven basis for determining the optimal number of clusters.

Employing k-means clustering alongside cosine similarity ensures that clusters represent distinct strategic scenarios accurately. This method enhances interpretability, allowing for clear delineation of game scenarios that consistently benefit one player or the other. The adoption of k-means thus supports an insightful and robust analysis of strategic patterns within the dataset.

2.4. Evaluation Metrics

Internal evaluation of the clustering results will be carried out using the silhouette index. This index measures how similar an individual move sequence is to its own cluster compared to other clusters, thereby indicating the quality of cluster separation. By computing the average silhouette score for different numbers of clusters, it becomes possible to identify the number of clusters that best captures the underlying structure in the data. A higher average silhouette score implies tighter, more cohesive clusters, guiding the selection of an optimal cluster count.

In addition to internal metrics, external evaluation will leverage resampling combined with the Adjusted Rand Index (ARI) [16], [17]. Since there is no definitive ground truth labeling for these generated sequences, the stability of clustering assignments under repeated sampling serves as a proxy for reliability. Specifically, multiple resampled subsets of the data will be clustered independently, and the ARI will compare cluster memberships across these iterations. A high ARI value indicates that members consistently group together despite variations in the dataset, signifying robust and stable clustering.

Together, the silhouette index and ARI-based resampling offer complementary perspectives: the silhouette index ensures that clusters are well-separated and cohesive, while ARI-based resampling verifies that these clusters remain consistent when data variations occur [18]. This dual approach ensures both the internal validity and external stability of the clustering results, strengthening confidence in the identified strategic patterns.

3. Results and Discussions

To explore strategic patterns in chess more deeply, this study conducted a clustering analysis on 5,156 chess games generated using the Stockfish engine version 17. The use of this engine ensured accurate and reliable evaluations of positions, as well as the generation of high-quality alternative move suggestions. The analysis specifically focused on the French Defense, aiming to understand its various lines and assess the potential advantages it offers to both Black and White. Through clustering, the study identified distinct patterns and similarities across different game paths. These findings provide the foundation for a further discussion on gameplay typologies and their strategic implications.

3.1. Clustering Results with Three Groups

Clustering analysis was carried out using the k-means algorithm with cosine similarity as the distance metric. Based on prior considerations and internal evaluation metrics, the number of clusters was set to 3. The goal of this clustering process was to categorize chess game sequences into three strategic outcome groups: favoring White, unclear, and favoring Black.

Table 1. The number of chess games included in each cluster and their percentages

Cluster Good for White	Cluster Unclear	Cluster Good for Black	Total
2518	900	1738	5156
48.84%	17.46%	33.70%	100%

As shown in Table 1, the clustering results grouped 2,518 sequences into the "favoring White" cluster, 900 sequences into the "unclear" cluster, and 1,738 sequences into the "favoring Black" cluster. This distribution suggests that a significant portion of the generated game sequences demonstrated positional advantages for White, as assessed by the Stockfish engine.

Examples of sequences that belong to the cluster favoring White are displayed in Figure 4. These include opening lines such as 1. exd5 a6 2. Nf3 Nf6 3. Bg5 Be7 4. dxe6 Rf8 5. c4 h6 6. Bf4 c5 7. Nc3 cxd4 8. exf7+, among others. The corresponding positional evaluation scores from the engine, shown in Figure 5, further support this categorization. For example, the evaluation scores for one such sequence show consistently high positive values such as 704, 524, 579, 408, and 428, indicating sustained positional advantage for White over a long series of moves. Another sequence displays scores like 511, 180, 336, 275, and 291, which also largely remain in positive territory.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
549	exd5	a6	Nf3	Nf6	Bg5	Be7	dxe6	Rf8	c4	h6	Bf4	c5	Nc3	cxd4	exf7+
356	e5	Nh6	c3	c5	Bxh6	Nc6	Be3	Bd7	a4	cxd4	Bf4	dxc3	Nf3	Rg8	bxc3
519	e5	Bd7	Nf3	c5	c4	Nc6	Bf4	h6	h4	dxc4	dxc5	Rc8	Nc3	Na5	Be2
56	exd5	exd5	Bd3	Bd6	Nc3	Ne7	Nb5	Bf5	Bg5	f6	Bxf5	Bb4+	c3	c6	Be6
124	Nc3	a6	Nf3	Nf6	Qd3	Bb4	Nd2	c5	a3	Ba5	exd5	Bxc3	bxc3	c4	Nxc4

Figure 4. Five examples of sequence of moves in chess games that falls into the first cluster, which favors White. From 1st move to the 15th.

X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30
267	224	254	245	226	184	188	155	168	166	147	117	131	125	192	198	189
308	315	365	367	358	351	357	295	295	263	348	308	386	399	430	436	518
72	26	67	41	42	30	160	195	222	91	176	171	320	285	232	234	270
349	277	332	347	529	495	704	524	579	408	428	391	404	320	404	350	489
95	96	91	90	511	180	336	-65	275	216	225	257	269	267	291	286	327

Figure 5. Five examples of the corresponding position evaluation scores of moves in chess games that falls into the first cluster, which favors White. From 14th move to the 30th.

Further analysis was conducted to evaluate the distribution of clusters across different opening variations. Table 2 presents the percentage of games for five move variations (e5, exd5, Nc3, Nd2, and Bb5+) categorized into the three clusters. The e5 variation, representing the Advance variation in French Defense, had the highest proportion of sequences (55.86%) falling into the favoring White cluster, followed by exd5 (54.56%). This suggests that these two variations are particularly promising for White. In contrast, the exd5 variation exhibited the lowest proportion of sequences categorized as favoring Black (26.99%), implying that this line tends to be less advantageous for Black.

Table 2. The percentages of chess games included in each cluster for each move variations for three clusters

Moves Variation	Cluster Good for White	Cluster Unclear	Cluster Good for Black
e5	55.86	16.70	27.44
exd5	54.56	18.45	26.99
Nc3	42.22	18.38	39.40
Nd2	45.47	17.43	37.10
Bb5+	44.96	16.38	38.66

Meanwhile, the Nc3, Nd2, and Bb5+ variations showed more similar distributions among the clusters. For instance, Nc3 had 42.22% of sequences favoring White, 18.38% unclear, and 39.40% favoring Black. Similarly, Nd2 showed 45.47% favoring White, 17.43% unclear, and 37.10% favoring Black. Bb5+ had 44.96% favoring White, 16.38% unclear, and 38.66% favoring Black. These results reinforce the strategic potential of the Advance and Exchange variations to tilt the position in favor of White more frequently than the other lines considered.

3.2. Clustering Results with Silhouette

To determine the most appropriate number of clusters, silhouette analysis was conducted as an internal validation method. The silhouette index measures how well each data point fits within its assigned cluster, with higher values indicating more coherent and well-separated clusters.

In this analysis, silhouette scores were calculated for cluster counts ranging from $k = 2$ to $k = 10$. The results, shown in Figure 6, indicate that the highest silhouette score (0.65) was achieved when the data was divided into two clusters. This suggests that the most optimal clustering structure, based on internal cohesion and separation, is with $k = 2$. However, the silhouette score for $k = 3$ was also reasonably high (above 0.5), which still falls within the acceptable range for good clustering performance.

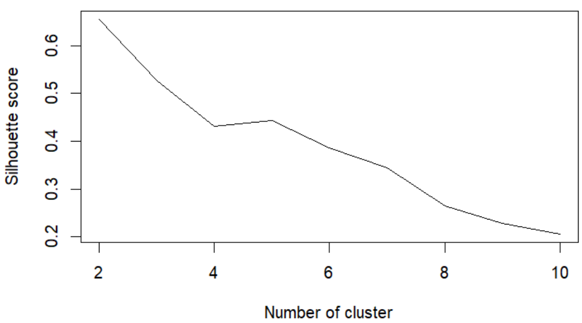

Figure 6. Plot of number of clusters against its silhouette score.

With $k = 2$, the model effectively partitions the dataset into two major categories: one favoring White and the other favoring Black. This configuration removes the intermediate or unclear cluster, meaning that sequences previously classified as unclear are now distributed into one of the two remaining groups based on proximity.

Table 3 shows how this reclassification impacts the interpretation of different opening variations. Notably, the e5 and exd5 variations still exhibit strong potential for White, each with 64% of their sequences placed in the favoring White cluster under the two-cluster model. In contrast, the Nc3, Nd2, and Bb5+ variations also lean towards White but with lower margins, making their advantage less pronounced compared to the first two. This distinction underlines the consistent strength of the Advance (e5) and Exchange (exd5) variations even when the classification scheme is simplified.

Table 3. The percentages of chess games included in each cluster for each move variations for two clusters

Moves Variation	Cluster Good for White	Cluster Good for Black
e5	64.21	35.79
exd5	64.37	35.63
Nc3	52.22	47.78
Nd2	54.63	45.37
Bb5+	52.7	47.3

3.3. Stability Evaluation of Clusters

To evaluate the robustness of the clustering results, a stability assessment was performed using a resampling technique, as in [19]. Resampling is commonly employed in machine learning and statistical analysis to test the consistency of model results under different subsets of data. It provides insight into whether the patterns discovered in the original dataset persist across randomly drawn samples, thus offering an empirical view of the clustering model's reliability.

In this study, stratified sampling was applied to each of the five major move variations (e5, exd5, Nc3, Nd2, and Bb5+). For each resample, 50% of the data from each variation was randomly selected, preserving the distribution of move types. The clustering process—using the same method and parameters as in the full dataset—was repeated 50 times on these resampled datasets, and the average results were recorded.

The outcomes for the three-cluster model are presented in Table 4. The average percentage of data points assigned to each cluster across the 50 runs closely mirrors the original proportions shown in Table 2. This comparability indicates that the overall structure and composition of the clusters are stable, even when only half of the data is used. Furthermore, the Adjusted Rand Index (ARI) was computed for each run to quantify the similarity between the clustering result of the resample and the original dataset. The average ARI score reached an impressive 0.99, indicating near-perfect alignment. A score this high suggests that the clustering assignments are highly consistent and minimally affected by data perturbation.

Table 4. The average percentages of chess games included in each cluster for each move variations for three clusters after 50 times resampling

Moves Variation	Cluster Good for White	Cluster Unclear	Cluster Good for Black
e5	55.66	16.80	27.54
exd5	54.24	18.71	27.05
Nc3	42.38	18.17	39.45
Nd2	45.56	17.33	37.11
Bb5+	44.68	16.46	38.86

Similarly, Table 5 reports the results for the two-cluster model under the same resampling conditions. The distribution of sequences into clusters remains consistent with those found in Table 3, again demonstrating stability in the proportion of favoring White versus favoring Black sequences. However, the average ARI in this case was lower, at 0.68. While this score still reflects a moderate to strong level of agreement, it also implies greater sensitivity to data variation when simplifying the classification to only two clusters. This is understandable, as forcing the previously "unclear" sequences into a binary structure can increase ambiguity and reduce alignment with the original clustering.

Table 5. The average percentages of chess games included in each cluster for each move variations for two clusters after 50 times resampling.

Moves Variation	Cluster Good for White	Cluster Good for Black
e5	63.85	36.15
exd5	64.05	35.95
Nc3	52.30	47.70
Nd2	54.50	45.50
Bb5+	52.41	47.59

3.5. Discussions

This study demonstrates that even a simple clustering method such as k-means can effectively uncover strategic patterns within chess games. Despite its simplicity, the method successfully grouped sequences in a meaningful way, revealing underlying structures and outcomes based on game variations.

The purpose of using cosine similarity with k-means clustering is to group chess games based on evaluation trends. Unlike Euclidean distance, which measures only the magnitude of differences, cosine similarity captures the direction of change—crucial for identifying whether a game favors White or Black. Dynamic Time Warping (DTW) is less suitable, as chess games vary widely and matching exact sequences is not the goal. Instead, we focus on overall trends in evaluation. Cosine similarity, paired with k-means, effectively group games with similar directional patterns.

The performance of the clustering approach is supported by both internal and external validation metrics. The silhouette scores—0.65 for the two-cluster model and 0.53 for the three-cluster model—indicate a good level of separation between clusters. Additionally, the stability of these models was evaluated through resampling. For the three-cluster model, the average Adjusted Rand Index (ARI) after 50 stratified resamples reached 0.99, reflecting high consistency and robustness[20]. Conversely, the two-cluster model, while achieving a higher silhouette score, yielded a lower ARI of 0.68, indicating a greater sensitivity to data variations.

This contrast in performance between the two models is notable. While the two-cluster model appears optimal from an internal cohesion perspective, its susceptibility to shifts in input data suggests that its simplicity may overlook structural ambiguities in the original dataset. A likely explanation is that removing the "unclear" cluster forces ambiguous sequences to align with either the White or Black favoring clusters. These ambiguous sequences can significantly shift the centroids of the remaining clusters, resulting in instability when small changes are introduced. This warrants deeper investigation into the behavior and influence of borderline data points.

From a gameplay perspective, the results are both surprising and insightful. The dataset, derived from Stockfish 17 engine simulations selecting the top five candidate moves per position, inherently assumes high-quality decision-making. However, in many positions, only one or two options are strong, with the remainder being suboptimal or blunders. This indicates that some opening lines, while seemingly reasonable, are risky due to a narrow margin for error, requiring extremely accurate play.

A particularly striking finding is the dominance of White in the e5 and exd5 variations. These two lines consistently fell into clusters favoring White, implying a significant strategic advantage when playing these variations within the French Defense. While it is widely acknowledged that White generally holds a slight initiative in chess, the extent of dominance shown in these simulations, especially for exd5, was unexpected.

To further interpret these results, Table 6 presents empirical game outcomes from the chess.com database. The data confirms that e5 and Nc3 yield relatively high win percentages for White (41% and 43%, respectively), consistent with the clustering results. Interestingly, exd5 in the real-world dataset is associated more with draws (52%) than outright wins for White (23%), suggesting a more balanced or drawish nature in practice. Nonetheless, the clustering analysis still highlighted exd5 as a safe and favorable path for White in the simulation, possibly due to the reduced likelihood of blunders within the early game phase.

Table 6. The winning percentage of the five-variation based on chess.com database[21]

Opening Variation	White Wins	Draw	Black Wins
e5	41	26	33
exd5	23	52	25
Nc3	43	29	28
Nd2	39	38	23
Bb5+	33	17	50

The Bb5+ variation also reveals an important discrepancy. In the simulation, it was grouped moderately with sequences favoring White. However, the real-world statistics indicate that 50% of games using this line result in a win for Black, making it the most dangerous option among the five. This mismatch may stem from the limitations of the simulated dataset, which only analyzes the first 15 moves of a game. In contrast, the outcomes in the chess.com database span the full game, including critical middle and endgame phases where imbalances may become more pronounced.

4. Conclusion

This study has demonstrated that even a simple clustering algorithm such as k-means, when combined with cosine similarity, can effectively uncover strategic patterns in chess games, particularly in the context of the French Defense. The use of cosine similarity proved to be a critical choice, as it captures not only the magnitude but also the direction of evaluation trends, which is essential in identifying whether a sequence of moves tends to favor White or Black.

The evaluation of clustering performance using both internal (silhouette index) and external (Adjusted Rand Index via resampling) metrics further reinforces the reliability of the findings. While the two-cluster model achieved a higher silhouette score (0.65), suggesting better cohesion and separation between clusters, the three-cluster model offered a more robust structure as shown by its high ARI score of 0.99. This indicates that including an “unclear” cluster provides greater stability against data variation, which is especially important in modeling ambiguous or balanced positions in chess.

Among the five variations analyzed, the e5 (Advance) and exd5 (Exchange) variations consistently showed the highest proportion of sequences falling into clusters favoring White. The results strongly support the conclusion that e5 is a safe and strategically advantageous choice for White within the French Defense, while exd5 also offers a reliable alternative that minimizes risk and yields favorable outcomes. This aligns partially with real-world game statistics, although some discrepancies—particularly with exd5's drawish nature in practical games—highlight the value of simulation-based evaluations in exposing latent advantages that may be obscured by human error in real matches.

On the other hand, Bb5+, although not part of traditional French Defense theory, showed potential merit in the simulations. However, its poor real-world performance—where it resulted in 50% wins for Black—calls for further theoretical investigation. Similarly, the Nc3 and Nd2 variations yielded more balanced clustering results, suggesting that while they are playable, their impact is more dependent on precise execution and may not consistently offer White a clear strategic edge.

Other important findings include the observation that some seemingly playable variations are highly volatile, requiring near-perfect play to maintain an advantage. This underscores the relevance of evaluating chess openings not just by theoretical soundness, but also by their practical margin for error—an area where simulation-based clustering can provide new insights.

Future research can explore deeper move sequences beyond the 30-move simulations used here, incorporate reinforcement learning agents to simulate more realistic decision trees, or extend the analysis to other openings and game phases. Additionally, the role of resampling in clustering evaluation should be further examined. Specifically, future studies could investigate how resampling impacts clustering evaluation indices and assess its potential in enhancing the validity and reliability of clustering outcomes.

Reference

[1] H. V. Ribeiro, R. S. Mendes, E. K. Lenzi, M. del Castillo-Mussot, and L. A. N. Amaral, "Move-by-Move Dynamics of the Advantage in Chess Matches Reveals Population-Level Learning of the Game," PLoS One, vol. 8, no. 1, pp. 1–7, 2013, doi: 10.1371/journal.pone.0054165.

[2] S. Djuric, D. Komarov, and C. Pantaleoni, Chess Opening Essentials. Alkmaar: New in Chess, 2007.

[3] T. Cook, "The Advantage of Moving First in Amateur Online Chess," ICGA Journal: The Journal of the Computer Games Community, vol. 46, no. 3–4, pp. 112–120, 2024, doi: 10.1177/13896911251315903.

[4] M. Barthelemy, "Space control and tipping points," pp. 1–10.

[5] A. Gupta, C. Grattoni, and A. Gupta, "Determining Chess Piece Values Using Machine Learning," Journal of Student Research, vol. 12, no. 1, pp. 1–20, 2023, doi: 10.47611/jsrhs.v12i1.4356.

[6] S. Chowdhary, I. Iacopini, and F. Battiston, "Quantifying human performance in chess," Sci Rep, vol. 13, no. 1, pp. 1–8, 2023, doi: 10.1038/s41598-023-27735-9.

[7] F. Wijayanto, "Forsyth-Edwards Notation in Chess Game Clustering: A Depth-Based Evaluation," Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi, vol. 4, no. 1, pp. 18–25, 2024, doi: 10.20885/snati.v4.i1.3.

[8] F. Wijayanto, "Clustering Analysis of Chess Portable Game Notation Text," Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi, vol. 3, no. 3, pp. 137–142, 2024, doi: 10.20885/snati.v3.i3.42.

[9] M. K. Khan, S. Sarker, S. M. Ahmed, and M. H. A. Khan, "K-Cosine-Means Clustering Algorithm," in Proceedings of International Conference on Electronics, Communications and Information Technology, ICECIT 2021, 2021. doi: 10.1109/ICECIT54077.2021.9641480.

[10] D. Dhillon, I. S. and Modha, "Concept Decompositions for Large Sparse Text Data.pdf," Mach Learn, vol. 42, pp. 143–175, 2001, [Online]. Available: https://link-springer-com.unimib.idm.oclc.org/content/pdf/10.1023/A:1007612920971.pdf?pdf=inline link

[11] G. De Marzo and V. D. P. Servedio, "Quantifying the complexity and similarity of chess openings using online chess community data," Sci Rep, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-31658-w.

[12] I. Cheng and I. Cheng, "Machine Learning to Study Patterns in Chess Games," University of Exeter, 2024. doi: 10.13140/RG.2.2.30894.52807.

[13] V. Mehta, S. Bawa, and J. Singh, "Analytical review of clustering techniques and proximity measures," Artif Intell Rev, vol. 53, no. 8, pp. 5995–6023, 2020, doi: 10.1007/s10462-020-09840-7.

[14] M. Rahul, P. Pal, V. Yadav, D. K. Dellwar, and S. Singh, "Impact of similarity measures in K-means clustering method used in movie recommender systems," IOP Conf Ser Mater Sci Eng, vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012101.

[15] A. Dudek, "Silhouette Index as Clustering Evaluation Tool," in Classification and Data Analysis, K. Jajuga, J. Batóg, and M. Walesiak, Eds., Cham: Springer International Publishing, 2020, pp. 19–33.

[16] S. Dudoit and J. Fridlyand, "A prediction-based resampling method for estimating the number of clusters in a dataset," Genome Biol, vol. 3, no. 7, pp. 1–21, 2002, doi: 10.1186/gb-2002-3-7-research0036.

[17] W. Wu, Z. Xu, G. Kou, and Y. Shi, "Decision-Making Support for the Evaluation of Clustering Algorithms Based on MCDM," Complexity, vol. 2020, 2020, doi: 10.1155/2020/9602526.

[18] C. Hennig, "Cluster-wise assessment of cluster stability," Comput Stat Data Anal, vol. 52, no. 1, pp. 258–271, 2007, doi: https://doi.org/10.1016/j.csda.2006.11.025.

[19] N. Meinshausen and P. Bühlmann, "Stability selection," J R Stat Soc Series B Stat Methodol, vol. 72, no. 4, pp. 417–473, 2010, doi: 10.1111/j.1467-9868.2010.00740.x.

[20] M. J. Warrens and H. van der Hoef, "Understanding the Adjusted Rand Index and Other Partition Comparison Indices Based on Counting Object Pairs," J Classif, vol. 39, no. 3, pp. 487–509, 2022, doi: 10.1007/s00357-022-09413-z.

[21] "Chess.com." [Online]. Available: https://www.chess.com