Chess Player Profiling: A Game - Changing Perspective on Player Development

Akanksha Funde[1], Prajwal Zade[2], Yash Khandebharad[3], Harsh Gharate[4]
[1,2,3,4]PES Modern College of Engineering, Savitribai Phule Pune University, India

***Abstract*— In the age of big data, significant quantities of different data can be fluently collected or generated at a rapid-fire pace. This data holds precious information that necessitates the use of machine literacy ways for advanced knowledge discovery. Game data, including data from sports games, card games, online videotape games, and chess games, serves as a precious source of big data. Chess, known for its profound commerce and straightforward representation, ranks among the most considerably delved games encyclopedically. multitudinous studies have been conducted in the once exercising chess data. These studies aimed to delve into the vast amount of chess data and employ machine learning models to classify it into various groups. Upon conducting our own exploration, we made an interesting discovery- a significant number of websites warrant comprehensive analysis for player development and are noticeably deficient in terms of coffers. In this paper, to address this issue, we propose the development of a dynamic dashboard with an expansive range of analysis and perceptivity, all grounded on the data they give. Our approach involves the perpetration of unsupervised literacy ways to construct a model able of grading players according to their scores. latterly, we will be suitable to offer acclimatized coffers and support to players grounded on their separate orders, with the ultimate end of fostering and enhancing their growth within the realm of chess.**

Keywords— machine learning, chess game, big data, data analysis, Django, database management.

I. INTRODUCTION

Chess is an incredibly demanding and intellectually stimulating sport that requires not only advanced planning and strategic thinking chops but also a deep understanding of the game's complications. The complexity and depth of chess make it a witching pursuit for suckers seeking internal challenges. One of the factors that adds to the appeal of chess is the time element. The duration of a game can vary significantly, ranging from bare twinkles to several knockouts of twinkles or indeed hours. This time factor not only affects the overall length of the game but also influences the intensity and pace at which it's played. It adds another subcaste of complexity and strategic decision- making for players to consider. To feed to the different preferences and playing styles of chess suckers, different formats have surfaced. These formats, similar as standard, blitz, and pellet, offer unique challenges and attract players with specific skill sets. In the standard format, players have further time to assay and plan their moves, while blitz and pellet formats demand quick thinking and fast decision-timber. Each format presents its own set of obstacles and requires players to acclimatize their strategies consequently.

The actuality of these distinct formats poses an intriguing dilemma when it comes to comparing players and determining the stylish chess player. Due to the technical nature of each format, it can be delicate to directly compare players from different orders. A player who excels in blitz chess may not inescapably perform as well in standard chess, and vice versa. This difference in chops across formats creates a fascinating challenge for chess suckers and experts likewise. Creating a unified ranking system that encompasses all types of chess, anyhow of the time constraints, is a significant challenge. numerous outstanding chess players specialize in specific formats and may not exceed in others.

Thus, chancing a fair and comprehensive result to rank players across all orders remains an ongoing bid in the world of chess. It requires careful consideration of colourful factors, including performance in different formats, thickness, and overall donation to the advancement of the game. Chess is a witching and intellectually stimulating sport that offers a myriad of challenges and openings for players to showcase their chops. The intricate nature of the game, coupled with the different formats and technical skill sets, adds complexity to the task of determining the stylish chess player.

Players not only contend by fighting for better positions but also by strategically exploring uncharted home in the game, which they believe their opponent may be less familiar with. This adds an redundant subcaste of complexity and excitement to the gameplay. To more understand and assay these strategic moves, multitudinous studies have employed supervised and unsupervised machine literacy ways. By applying these ways, experimenters are suitable to classify analogous moves in games and indeed prognosticate unborn game data grounded on a model created using literal game data. This not only enhances our understanding of the game but also provides precious perceptivity for players to ameliorate their own strategies.

Likewise, there have been other notable exploration sweats that concentrate on using underpinning learning styles to automatically design an AI player that's amusing for mortal players, especially those who may not exceed at playing games. This innovative approach involves the AI player learning through playing against a computer opponent, rather than a mortal player. By doing so, the AI player can acclimatize and learn from its gests to produce a more engaging and gruelling gameplay experience. It's important to punctuate that while AlphaGo Zero employed the AI player as its own opponent, this approach may not be suitable for the proposed system. This is because the AI player in the proposed system serves a different purpose compared to that of a mortal player. nevertheless, the eventuality of using underpinning learning ways to enhance the gaming experience for players is a promising avenue for unborn exploration and development.

Although players are ranked grounded on being standing systems, the use of robotization plays a vital part in efficiently coordinating matches in events with a large number of actors. To assign conditions to players without being conditions, several studies have conducted detailed analyses of game records of largely professed chess players. These analyses have employed a decision tree model to determine applicable conditions for unrated players. The results of these analyses have consistently demonstrated the ability to accurately identify players as high-rated or low-rated with accuracy levels surpassing 80%. This holds true even for players who were not part of the model's training.

II. BASIC METHODOLOGY

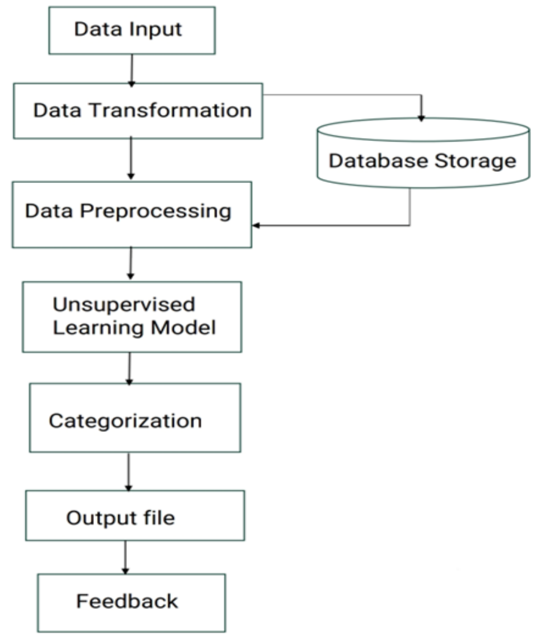

Fig. 1. Chess data analysis stages

A. Data Input

The data, provided by the user, is stored in a Portable Game Notation (PGN) file format. This file format is commonly used to store chess game information. The PGN file contains a wealth of information related to the chess game, ensuring that all the important details are captured and preserved. In addition to the moves played by the player, the PGN file also includes various other valuable details. These details encompass the tournament name, location, and date, allowing for easy identification and categorization of the game. Furthermore, the PGN file provides important information about the players involved, including their names and respective ratings. This not only adds depth to the game record, but it also enables easy tracking and analysis of player performance. Last but not least, the PGN file also records the results of the games, ensuring that the outcome is accurately documented.

B. Data Transformation

The conversion process itself is designed to ensure data integrity and accuracy. Through careful parsing and transformation techniques, we are able to preserve the essential information contained in the PGN file while organizing it in a structured and standardized manner within the csv file. This systematic approach not only improves data quality and consistency but also enables seamless integration with other data sources and applications, fostering a comprehensive and interconnected data ecosystem. Additionally, the csv format allows for compatibility with a wide range of data analysis tools and software, expanding the possibilities for in-depth exploration and interpretation of the chess data.

C. Data Preprocessing

In the next stage of data preprocessing, various methods are performed to ensure the quality of the data. These methods include data cleaning, where any inconsistencies or errors in the data are identified and resolved. Additionally, outliers are analyzed to determine if they should be included in the dataset or treated separately. Feature extraction is another important step, where relevant features are selected or created from the existing data to improve the model's performance. All these steps are crucial in providing the model with accurate and meaningful data, which in turn helps in training the model effectively.

D. Unsupervised Learning Model

An unsupervised learning algorithm, such as K-means clustering, can be employed to assess the value of data on a scale of 0 to 100. This powerful algorithm categorizes the data into distinct levels of expertise, encompassing novices (0-25), those with intermediate skills (25-50), individuals with advanced capabilities (50-75), and true masters of the field (75-100). This classification is based on a comprehensive analysis of diverse features and characteristics. By leveraging the capabilities of this algorithm, we are able to delve deeper into the intricacies of the data and proficiently classify it into its respective proficiency levels.

E. Output File

After classifying and organizing the data into various distinct groups based on the model, we will ensure that each group receives an extensive array of resources pertaining to the game of chess. These resources will be meticulously customized to meet the specific requirements and preferences of each group. Among the resources provided will be comprehensive tutorials that cover all aspects of chess, in-depth analyses of renowned chess matches to enhance understanding, practical exercises designed to enhance strategic thinking abilities, and recommendations for further study to deepen knowledge. By offering such a diverse and specialized range of resources, our objective is to cater to the unique needs and interests of every individual within the categorized groups, thus guaranteeing a comprehensive and highly enriching learning experience in the fascinating realm of chess.

F. Feedback

Feedback also plays a vital role in fostering collaboration and engagement. It encourages open communication between our team and our users, creating a dynamic and interactive environment where ideas can be exchanged and knowledge can be shared. Through this collaborative feedback loop, we can continuously evolve and adapt, staying at the forefront of advancements in data analysis and ensuring that our users receive the highest quality support and resources.

III. LITERATURE SURVEY

In 2017, Roopali Garg and Deva Prasad Nayak introduced a simulation algorithm to predict the outcome of a tic-tac-toe game based on the player's first move. This algorithm utilizes the Min-max algorithm and combinational game theory to determine the best move, increasing the player's chances of winning and decreasing the opponent's chances. It assesses all possible moves and their potential outcomes to create a strategy that cannot be beaten. Since tic-tac-toe has a small 3x3 grid, the algorithm can thoroughly evaluate all moves and their outcomes. By considering possibilities and potential outcomes, the algorithm ensures that the player makes advantageous moves at each step, with the goal of winning the game and maximizing profit by creating forks or securing wins. These principles and techniques can be applied to more complex games like chess and checkers, where evaluating moves and predicting outcomes are crucial. Further research can expand upon this algorithm to develop unbeatable strategies for these games. By utilizing the Min-max algorithm and combinational game theory, the algorithm maximizes the player's chances of winning while minimizing the opponent's chances [1]. It serves as a foundation for developing unbeatable strategies in chess and checkers as well. Overall, it demonstrates the potential of game theory and algorithms in enhancing game-playing strategies and decision-making.

In 2017, James A. Brown and his co-operative researchers proposed a machine learning system that uses unsupervised and supervised learning components to analyze a large amount of chess data. The evaluation results [2] show that machine learning not only helps find interesting games but also demonstrates that chess can be a valuable testing ground for machine learning and data mining techniques in big data analytics. Unsupervised learning, also known as clustering, aims to organize unlabeled data into distinct groups or clusters based on their similarity. The resulting clusters can provide insights into the distribution of data or serve as input for supervised learning tasks. Some common approaches to unsupervised learning include:1. Partitioning approaches: These create k partitions represented by k means or k medoids and then assign the remaining objects into these partitions based on their similarity distance. 2. Hierarchical approaches: These either start with individual objects and merge similar ones to form clusters in a bottom-up fashion or start with a single group containing all objects and split it into smaller clusters in a top-down fashion, both based on similarity distance. Supervised learning, on the other hand, focuses on building a classification model using historical training data and then using that model to predict class labels for future data. Common approaches to supervised learning include decision trees, Bayesian belief networks, neural networks, support vector machines (SVM), and associative classification. As part of ongoing work, they are conducting more comprehensive evaluations by comparing different machine learning and data mining approaches to support advanced knowledge discovery from big chess game data.

In 2018, Zhi Wang and Xizhao Wang proposed a set of chess situation features based on their specialized knowledge of Chinese chess. These features have been proven to be effective in estimating the situation. Another noteworthy aspect of this paper is the use of L1/2 regularization to constrain the autoencoder. This regularization technique increases the sparsity of the dataset, making the situation features more representative and mitigating overfitting issues with SWAN. The hidden layers in this model act as an autoencoder, while the last hidden layer serves as a neural network classifier trained using stochastic weight assignment network (SWAN). SWAN is an emerging algorithm for training single-hidden layer feedforward neural networks (SLFNs) [3]. The core idea behind SWAN is to randomly generate weights and biases between the input and hidden layers, and then analytically compute the weights of the output layer by solving a generalized inverse of the matrix. The authors demonstrated that the Chinese chess situation dataset is representative and addresses the lack of relevant Chinese chess data. They proposed a novel MLP architecture that combines L1/2 sparsity-constrained AE and SWAN classifier, called DSWAN. DSWAN is capable of extracting features and performing classification on the Chinese chess situation dataset. Due to the high dimensionality of the Chinese chess situation data, overfitting becomes a challenge during the learning process. However, by incorporating L1/2 regularization in the AE, the mapped features of the Chinese chess situation dataset become sparser. This effectively reduces the dimensionality before the SWAN classification phase, thus alleviating overfitting problems.

In 2020, Meiyang Li and Wenzhi Huang proposed a Chinese chess game algorithm based on reinforcement learning. The algorithm utilizes a self-play learning model that combines deep convolutional neural networks and the Monte Carlo search tree algorithm. It uses the Monte Carlo search tree algorithm [4] to simulate chess moves, allowing for constantly changing roles. The deep neural network enhances the chess strategy and is trained in the reverse direction after winning or losing, aiming to achieve a certain level of chess skill in Chinese chess through self-learning without initial training data. After a period of training, the model's self-learning ability was found to be strong, and it exhibited good performance in Chinese chess. The Monte Carlo algorithm can be applied to board games through its use. In the reinforcement learning model discussed in this article, the Monte Carlo algorithm differs from the traditional Monte Carlo algorithm in two aspects. First, the traditional MCTS selects the next step, while in this model, choices are made using two parameters, Q and U, initialized for all nodes. The reinforcement learning model shows potential in enhancing chess skills and demonstrates considerable self-learning ability.

In 2020, Siwen Pan, Jiehong Wu, Yanan Sun, and Yi Qi confidently proposed an algorithm to expand the game tree by updating the probability table based on the rules of military chess. This algorithm takes into account factors such as the opponent's moves and the collision of chess pieces. Military chess is an incomplete information game where two players, red and black, each have 25 chess pieces. The algorithm updates the probability table to provide scores for the expanded game tree, ensuring its growth[5]. The initial values and weights in the algorithm are optimized using machine learning to improve speed and accuracy. The game tree is further optimized using the minimax algorithm and the Alpha-beta pruning algorithm. Experimental results demonstrate that this algorithm significantly improves the winning rate in the game. The evaluation score of the game tree node consists of two parts: the score of chess piece movements, which represents the immediate benefit, and the situation evaluation score, which indicates the overall profit or loss after simulating chess movements. The algorithm for updating the probability table is not limited to military chess but can be applied to any incomplete information game to gradually gather information and make optimal decisions. While computer games have already reached their peak in terms of complete information games, future efforts will focus on developing technology and algorithms for incomplete information games. Obtaining more game information through different algorithms will be a crucial direction for their development.

In 2020, Takanobu Yaguchi and Hitoshi Iima proposed a confident method that utilizes artificial intelligence (AI) to create a computer opponent capable of entertaining players with limited gaming skills. The focus of this method lies in one-on-one games, with the aim of designing a computer opponent that strikes a balance between being formidable and being weak. The effectiveness of this method was thoroughly evaluated through numerical experiments. Within the game, two players take turns executing actions and earn scores based on their choices. The player with the highest score emerges victorious upon the game's conclusion. This game is deterministic in nature, devoid of randomness, and ensures both players possess complete knowledge about its mechanics[6]. While Q-learning, a renowned reinforcement learning technique employing state-action values, is commonly utilized, it can be time-consuming to learn the optimal actions in games with a vast number of states and actions. To expedite the learning process, the proposed method adopts the temporal difference (TD) method, which utilizes state values. In this approach, the AI player selects an action in a given state, observes the subsequent state, and is rewarded accordingly. The performance of the proposed AI player was evaluated by pitting it against 121 different computer players instead of human players who struggle with games. However, it is imperative to further evaluate the AI player against human opponents in future experiments. Furthermore, the concept behind this method could potentially be applied to deep reinforcement learning, a technique renowned for its efficacy in handling a multitude of states.

In 2020, Cundong Tang, Zhiping Wang, Xiuxiu Sima, and Lingxizo Zhang proposed a transformative model that redefined game development. Their proposal analyzed AI's historical context and its current role, while also predicting the future impact of machine learning-based AI on the industry. By utilizing AI, game developers unlocked new levels of player interaction and realism, transcending traditional programming constraints. However, challenges arose when describing complex non-player character behaviors, prompting a shift to machine learning solutions. This approach allowed developers to decode player behavior and enhance games continually. Moreover, the model introduced self-learning capabilities, enabling offline learning and parameter extraction for neural networks[7]. This fusion of human intuition and machine learning promised to revolutionize game development, paving the way for innovative game modes, gameplay dynamics, and overall player experiences. In essence, Tang, Wang, Sima, and Zhang's model heralded a thrilling new era in gaming, marked by limitless possibilities and unparalleled creativity.

In 2021, Rafał Dreżewski and Grzegorz Wątor conducted a rigorous investigation into the predictive capabilities of Recursive Neural Networks, particularly LSTM (Long Short-Term Memory) [8], for determining chess game outcomes. Their research encompassed the examination of various chess position representations, including bitmap and algebraic inputs, and introduced a neural network architecture that incorporated essential metadata about the participating players. In summary, the network model they devised comprised two integral components:1. A neural network that assessed the game's outcome based on data directly linked to the participating chess players. 2. A neural network that predicted the game's result based on the progression of the game itself. Their compelling findings demonstrated that longer sequences of gameplay data significantly improved prediction accuracy, with games spanning 80 moves achieving an impressive 65.32% accuracy rate, highlighting the profound influence of sequence length on the precision of chess game outcome predictions.

In 2021, Aleksandra Kaczyńska, Joanna Kołodziejczyk, and Wojciech Sałabun aimed to create a universal ranking system for chess players, regardless of the specific type or time control of the games they specialized in. Establishing a single ranking for just one type of chess wouldn't accurately represent players' overall skills, as many excel in different formats. The data of the evaluated chess players should fall into ranges: from 2736 to 2862 Classic ranking criterion [9]; from 2688 to 2863 Mean of classic ranking criterion; from 2683.5 to 2881 Rapid ranking criterion; from 265-4 to 2900 Blitz ranking criterion; from 18 to 39 Player's age criterion. In addition, average values for a given group of chess players were determined: 2782 Classic ranking criterion: 2767.6 Mean of classic ranking criterion; 2778.6 Rapid ranking criterion: 2776.3 Blitz ranking criterion; 31.2 Player's age criterion. To overcome this challenge, they employed the COMET method, a multi-criteria decision-making approach that uses fuzzy logic and expert input to assess players. To ensure reliability, they defined specific rating ranges for each chess format, such as classic, rapid, and blitz. This method proved effective in ranking chess players across various game types and could be useful for similar complex ranking problems in sports and beyond.

In 2023, Dimitrios Novas, Dimitrios Papakyriakopoulos, Elizabeth Powlesland Kartaloglou, and Anastasia Griva proposed a ranking model that extracts topics from qualitative data (documents) and converts them into quantitative values using fuzzy logic. Our model uses these topics as the ranking space and performs pairwise comparisons between items (restaurants) to create a ranking ladder. They applied model to analyze TripAdvisor's documents from restaurants in Athens, Greece [10]. It is important to note that the ranking of restaurants can vary significantly between platforms. For instance, TripAdvisor (2020) states that their ranking is based on the Popularity Index, which considers factors such as Quality (ratings from 1 to 5), Recency (latest comments), and Quantity (volume of comments). This research suggests that similar logic could be applied to other domains like hotels or airlines. Additionally, future studies could address the limitations of our research, such as the challenge of making direct comparisons with TripAdvisor's ranking model due to the lack of explainability of their platform. To overcome this, it would be beneficial to establish a benchmark ranking model that approximates the ground truth. Furthermore, further research could focus on evaluating and applying our model to other domains (e.g., hotels, movies, products) or using data from different platforms (e.g., Yelp). In conclusion, our study also highlights the importance of examining the ethical aspects of ranking systems, such as the fairness of existing systems and opportunities for improvement.

In 2023, Haize Hua, Jianxun Liua, Xiangping Zhanga, Mengge Fang proposed a new k-means clustering algorithm called Lévy flight trajectory-based k-means (Lk-means). The Lk-means algorithm uses Lévy flight during the iterative process to search for new positions, preventing premature convergence in clustering. This technique enhances cluster diversity, improves the global search ability of the K-means algorithm, and avoids early convergence to local optimal values. Importantly, the complexity of the hybrid algorithm remains unchanged during the Lévy flight optimization. To evaluate the data clustering performance of the Lk-means algorithm [11], experiments were conducted using 10 open source datasets. The algorithm was compared with other k-means algorithms, including XK-means, DDK-means, and Canopyk-means. The results demonstrate that the Lk-means algorithm achieves better search results and more evenly distributed cluster centroids. This greatly enhances the global search ability, big data processing capability, and uneven distribution of centroids in the K-means algorithm. It is worth noting that our comparative analysis only focused on k-means algorithms of the same type. Future research should include the evaluation of other types of algorithms, such as particle swarm optimization, Cuckoo Search, Ant Colony optimization, and machine learning models. These additional evaluations will further validate the effectiveness of the algorithm by continuously optimizing the clustering effect and exploring multiple types of comparative models.

IV. CONCULSION

With the ever-increasing popularity of chess, there has been a remarkable surge in both the production and accumulation of vast amounts of diverse chess data.

This extensive data encompasses a wide range of information pertaining to chess games and is being amassed at an astonishing pace. In this comprehensive research paper, we have introduced a cutting-edge machine learning system that aims to not only assist users in meticulously analyzing their game performance but also in evaluating their level of expertise. Moreover, our innovative system offers meticulously curated resources that are specifically tailored to cater to the unique requirements of individual users. These invaluable resources are intelligently generated based on the carefully determined categories derived from the model, which is meticulously constructed utilizing the PGN file provided by the user. As we continue to diligently work on advancing this groundbreaking project, we are actively planning to integrate even more sophisticated machine learning techniques to further optimize and enhance the overall accuracy and precision of the output categorization.

REFERENCE

[1] Roopali Garg and Deva Prasad Nayak, "GAME OF TIC-TAC-TOE: SIMULATION USING MIN-MAX ALGORITHM," *International Journal of Advanced Research in Computer Science.*, vol. 8, no. 7, pp. 1074-1077, Aug. 2017, doi: 10.26483/ijarcs.v8i7.4409

[2] James A Brown, Alfredo Cuzzocrea, Michael Kresta, Korbin D.L. Kristjanson C.K. Leung, T.W. Tebinka, "A Machine Learning System for Supporting Advanced Knowledge Discovery from Chess Game Data," *2017 16th IEEE International Conference on Machine Learning and Application.*, pp. 649–654, 2017.

[3] Zhi Wang, Xizhao Wang, "A deep stochastic weight assignment network and its application to chess playing," *J. Parallel Distrib. Comput.* , pp. 205–211, 2018.

[4] Meiyang Li, Wenzhi Huang, "Research and Implementation of Chinese Chess Game Algorithm Based on Reinforcement Learning,"*5th International Conference on Control, Robotics and Cybernetics*, pp. 81-86, 2020 IEEE.

[5] Siwen Pan, Jiehong Wu, Yanan Sun , Yi Qu, "A Military Chess Game Tree Algorithm Based on Refresh Probability Table," *IEEE International Conference.* , 978-1-7281-5855-6/20., pp. 588–591, 2020

[6] Takanobu Yaguchi, Hitoshi Iima, "Design of an Artificial Game Entertainer by Reinforcement Learning ," *IEEE International Conference.* , 978-1-7281-4533-4/20, pp. 588–591, 2020

[7] Cundong Tang, Zhiping Wang , Xiuxiu Sima , Lingxiao Zhang, "Research on Artificial Intelligence Algorithm and Its Application in Games," *2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM).*, 78-1-7281-9986-3/20/, pp. 386–389, 2020.

[8] Rafał Drezewski, Grzegorz Wątor, "Chess as Sequential Data in a Chess Match Outcome Prediction Using Deep Learning with Various Chessboard Representations," *25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems.*, pp. 1760–1769, 2021.

[9] Aleksandra Kaczyńska. Joanną Kolodziejczyk". Wojciech Satabun, "A new multi-criteria model for ranking chess players," *25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems.*, pp. 4291–4299, 2021.

[10] Dimitrios Novas , Dimitrios Papakyriakopoulos , Elizabeth Powlesland Kartaloglou, Anastasia Griva, "A ranking model based on user generated content and fuzzy logic," *International Journal of Hospitality Management.*, pp. 1–13, 2023.

[11] Haize Hu, Jianxun Liu, Xiangping Zhang, Mengge Fang, "An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis,"*Pattern Recognition.0031-3203*, pp. 1–18, 2023.