Findings Provided by a Machine Learning Clustering Model of Stellar Kinematics in Star Clusters Can be Used to Identify Runaway Stars

Roneet Dhar, Tony Rodriguez

ABSTRACT

Star clusters dissipate as internal and external gravitational interactions weaken its tidal radius, causing smaller groups of stars to split apart in several directions while pushing other stars along with it. The need for developing a machine learning model that can detect these runaway stars is established in this research as it attempts to expand the domain of astrophysics to help astrophysicists understand tidal tail formation and the theoretical dark matter subhalos' effects on celestial bodies. Data from the ESA's Gaia DR 3 was utilized and noise was removed through data cleaning with 4 parameters: RAdeg, DEdeg, PM, and Distance that were then factored into a K-means clustering algorithm, which created the star clusters. These 4 factors were subsequently compared and analyzed in various scatterplots. The findings revealed the model's success in identifying potential candidates of runaway stars by confirming the existence of high proper-motion stars amongst other more similar proper-motion stars and the ability for stars to be separate by vast distances across their cluster, indicating possible tidal tails. Future improvements to work around various limitations encountered throughout this research focus on the development of a hybrid clustering-nueral network model to automatically detect and predict the likelihood of runaway stars from their clusters.

INTRODUCTION

Star clusters are groups of 10 or more stars that were formed from the same interstellar cloud and are gravitationally bound to each other (10). The three types of star clusters are globular clusters, open clusters, and embedded clusters (10).

Globular Clusters

Globular clusters were formed in the beginning of the universe, as they require gas and dust from giant molecular clouds. They currently have exhausted most of their interstellar gas and dust supply and new ones are less common in the universe. Globular clusters are spherical groups of ancient stars that can contain millions of stars densely packed together around a central mass. The star motions are determined by the sum of the mass of all stars, as well as their "relaxation times," which is the time it takes for stars to become disturbed by the gravitational interactions with other stars around it. In globular clusters, star distribution is isotropic, and their orbits are virtually random (2). This results in the globular cluster's spherical shape, with their individual motions preventing them from crashing into each other, while gravity ensures that the "cluster does not fly apart" (3).

Although globular clusters are less likely to collapse, certain internal and external forces can rip some stars out of a globular cluster. Overtime, stars in a globular cluster will get close enough to interact gravitationally with each other. If two stars with starkly contrasting masses come close to each other, with one star being faster than the other, the faster star will donate energy to the slower star. The slower star in turn will gain a large amount of velocity and can escape the cluster if in the right position. This also implies that massive stars slow down over time, allowing gravity to pull them toward the center. As a result, the cluster segregates its stars by mass, with the heavier stars in the center and the lighter faster stars on the outskirts (9). If these stars are outside of the cluster's tidal radius, then they can be stripped away from the cluster (12).

Open and Embedded Clusters

Unlike globular clusters, open clusters are often irregular in shape, containing only dozens or hundreds of stars. Since open clusters are generally located closer to the center of the galactic disk, celestial bodies can easily pull them apart over long periods of time. As a result, they are typically younger than globular clusters, and reveal numerous datapoints that can be used for understanding star evolution. Meanwhile, embedded clusters are small groups of stars surrounded by interstellar gas and dust. These are often newly born or are forming, producing intense radiation. The current theory is that most stars originate from embedded clusters, and then transform to group into open clusters and beyond (10).

The Dataset (ESA Gaia DR 3)

This research helps to improve our understanding of the formation of star clusters and galaxies in our universe. It has several implications, including possible evidence of dark matter sub halos, the improvement of n-body time simulations showing the evolution of star clusters, and demonstrate the formation of tidal tails. The objective was to build a hybrid machine learning model using data from the ESA's Gaia Data Release (DR) 3 that could map the kinematics of stars in their clusters accurately and then predict which stars would leave their clusters. The primary variables utilized from the dataset were the right ascension degree (RAdeg), declination degree (DEdeg), proper motion (PM), and distance (1/Parallax). The RAdeg and DEdeg gave the positions (coordinates on the sky) of stars within their clusters, the proper motion gave the velocity of the stars across the sky, and the distance provided would help to identify the entire cluster's location in the sky, as well as the specific stars being analyzed.

MATERIALS & METHODS

In this study, the Gaia DR 3 40-parsec and 100-parsec datasets were utilized and their findings were compared. Noise from the datasets was reduced by removing all stars that had an RUWE index greater than 1.4 (to ensure stars in the dataset did not have faulty parallax or PM readings). In addition, data cleaning was applied to primarily focus on the 4 primary parameters for determining cluster membership: right ascension degree (RAdeg), declination degree (DEdeg), proper motion (PM), and distance (Dist), which was calculated by taking the reciprocal of the Gaia DR 3 parallax readings. These 4 parameters were then factored into an unsupervised K-means clustering algorithm to form the appropriate and optimal number of star clusters based on the heuristic elbow method.

The Machine Learning (ML) Pipeline

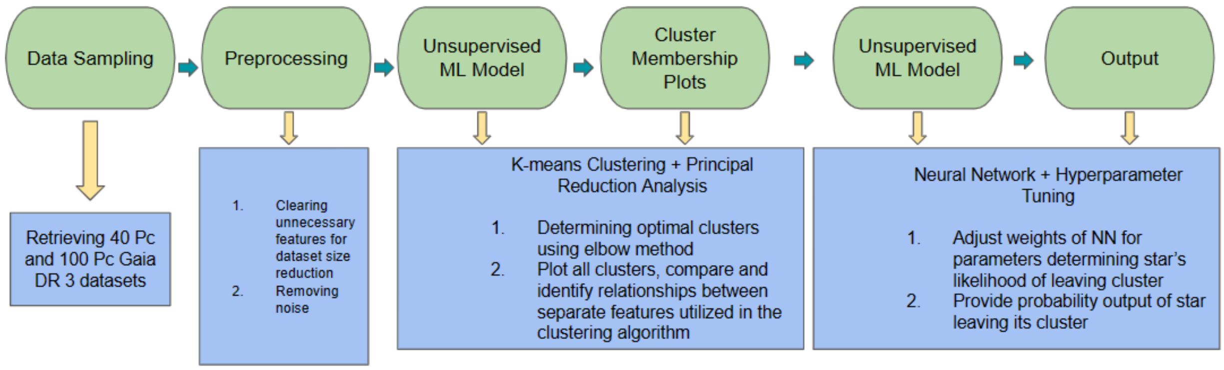


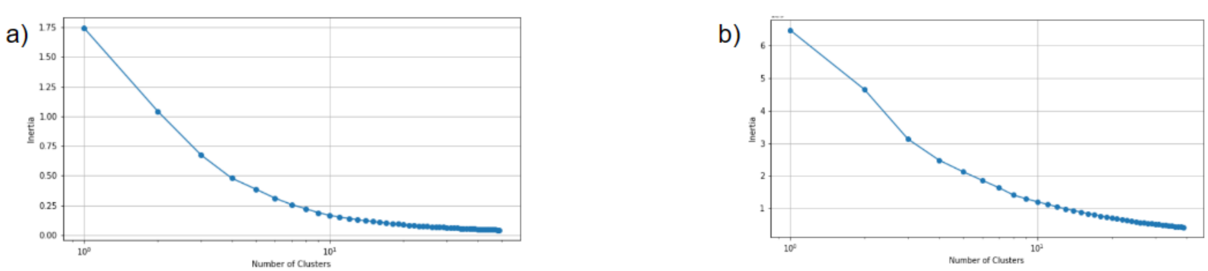
Figure 1: The **Machine Learning (ML) Pipeline** above visualizes the flow of the entire research design.

In order to determine the stars that could potentially leave their clusters, a sophisticated procedure was followed. First, the stars needed to be accurately clustered and the tidal radii of the star clusters needed to be calculated. Then, gravitational interactions with neighoring stars would be factored in, implementing n-body simulations in a neural network to predict each star's path after a gravitational interaction. If a star is on the edge of the cluster's tidal radius, has low mass, and has a high proper motion compared to its neighboring stars, then it is a likely candidate for leaving its cluster. Therefore, it would be assigned a high probability value for leaving its cluster (stars that fulfill less of these deterministic factors get assigned lower probability values for leaving their clusters).

RESULTS

Cluster Metrics

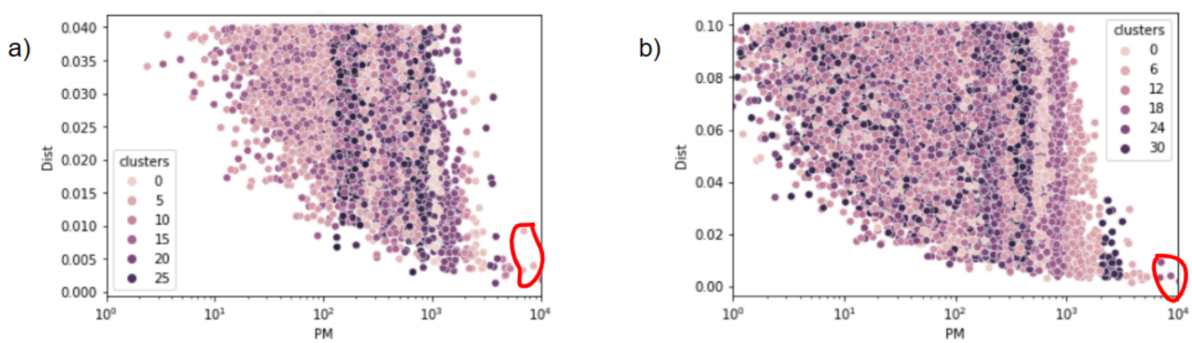
First, the optimal number of star clusters was calculated using elbow method (depicted below in Figures 2a-2b). The curve seems to decrease linearly (elbow) upon reaching n=30 clusters for the 40 parsec dataset and at n=36 clusters for the 100 parsec dataset; hence, these were the n-values for the k-means clustering algorithm for each respective dataset.



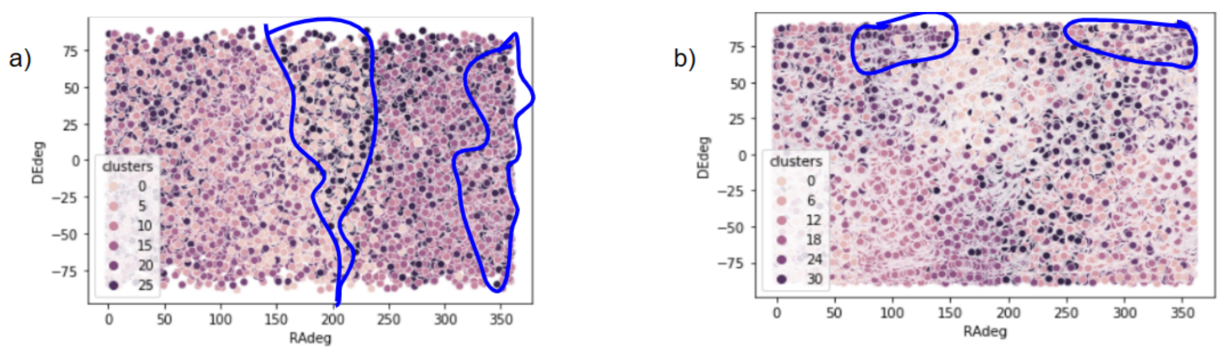
Figures 2a-2b: The **Elbow Graphs** provide the optimal number of clusters for each dataset (logarithmically scaled). Figure 2a outputs the optimal number of clusters in the 40 parsec dataset, while Figure 2b returns the optimal number of clusters in the 100 parsec dataset.

Cluster Plots

Upon running the K-means clustering algorithm on the datasets, each star cluster's proper motion vs. distance and RAdeg vs. DEdeg graphs were graphed (Figures 3a-3b, 4a-4b). These graphs revealed key findings in building a machine learning model to predict the likelihood of stars leaving their clusters. The Proper Motion vs. Distance graphs confirmed the existence of abnormal fast-moving stars in the star clusters, while the RAdeg vs. DEdeg graphs demonstrated "clusters of stars within star clusters" as they were separated by vast distances across the sky coordinates. These findings are discussed in more detail in the following section.



Figures 3a-3b: The **PM vs. Dist Cluster Plots** show the relationship between the proper motion and distance of stars of the same cluster with respect to their proper motions (logarithmically scaled). Figure 3a exhibits this relationship with the 40 parsec dataset, while Figure 3b displays the 100 parsec dataset. Key findings are indicated by the red circles.



Figures 4a-4b: The **RAdeg vs. DEdeg Plots** provide a map of stars' locations in their clusters. Figure 4a plots the positions of stars in their clusters for the 40 parsec dataset, while Figure 4b delineates star clusters in the 100 parsec dataset. Key findings are marked by the blue annotations.

Unfortunately, due to limitations (discussed in the next section), a neural network was unable to be implemented in order to fully determine stars that could leave their clusters. However, the plots above (Figures 3a-3b, 4a-4b) present valuable findings that can be used to build machine learning models that can automatically detect stars leaving their clusters.

DISCUSSION

Key Findings & Implications

As stated in the previous section, Figures 3a, 3b, 4a, and 4b yield findings with key implications: they confirm the existence of high proper-motion stars in the same cluster as other more similar proper-motion stars and the fact that stars can be separated by location by large distances in smaller groups, but can be still part of the same cluster (of any type such as globular clusters, open clusters, etc.). Stars with higher proper motions than others in their cluster can be ejected from their clusters if far enough from the cluster's center (as the tidal force won't be as strong to keep these stars within the cluster). Also, the fact that stars of the same cluster can be separated apart by relatively large distances, and be part of "mini-clusters" within their clusters, confirm that tidal tails are forming (most likely on opposite ends of the cluster), with the star cluster dissipating as portrayed in Figures 4a-4b. All of those stars in the tidal tail would therefore be likely candidates for leaving their cluster. A neural network can factor in all these determining factors by having appropriately adjusted weights to assign accurate likelihoods accordingly.

Limitations

However, several limitations occurred throughout the research process. A neural network was unable to be implemented due to time constraints and knowledge barriers, as well as a dataset with limited variables. In addition, the K-means clustering algorithm utilized in this paper could be improved to remove additional noise present in the dataset, and to ensure that the model's predicted clusters have accurate sizes (to ensure that certain stars are not missed or not wrongly counted as part of a cluster). Utilizing more sophisticated unsupervised clustering models such as DBScan and Guassian Mixture Models to complement a neural network (to form a hybrid model) would help refine the model's overall accuracy.

Next Steps

Ultimately, a model for determining stars that leave their clusters would never be fully accurate due to the n-body problem in physics. This unsolved problem states that due to several internal and external factors, it is impossible to predict the result of three or more celestial bodies gravitationally interacting. Current models only serve as approximations, and therefore a machine learning model would never be able to certainly predict the properties of stars (whether they become more or less massive, gain or lose speed, etc.) after these interactions. However, this paper attempts to inspire future attempts that strive to develop machine learning models that can identify stars that can leave their clusters with as much accuracy as possible.

REFERENCES

1. Castro-Ginard, A., et al. “Hunting for Open Clusters in *Gaia* EDR3: 628 New Open Clusters Found with Ocfinder.” *Astronomy & Astrophysics*, vol. 661, 24 May 2022, doi:10.1051/0004-6361/202142568.
2. Haskins, George. “What Are the Motions of Stars Relative to Each Other in a Globular Or.” *Astronomy.com*, Astronomy, 3 Jan. 2018, astronomy.com/magazine/ask-astro/2018/01/celestial-motion.
3. Howell, Elizabeth. “Globular Clusters: Dense Groups of Stars.” *Space.com*, Space, 22 July 2015, www.space.com/29717-globular-clusters.html.
4. KÃ¼pper, Andreas H., et al. “Tidal Tails of Star Clusters.” *Monthly Notices of the Royal Astronomical Society*, vol. 401, no. 1, Jan. 2010, pp. 105–120., doi:10.1111/j.1365-2966.2009.15690.x.
5. Liu, Lei, and Xiaoying Pang. “A Catalog of Newly Identified Star Clusters in *Gaia* DR2.” *The Astrophysical Journal Supplement Series*, vol. 245, no. 2, 12 Dec. 2019, p. 32., doi:10.3847/1538-4365/ab530a.
6. McDivitt, Jessica, et al. “Using Stellar Rotation to Identify Tidally Stripped Members of the Praesepe Open Cluster.” *ArXiv.org*, ArXiv, 10 June 2022, arxiv.org/abs/2206.05179.
7. Pang, Xiaoying, et al. “3D Morphology of Open Clusters in the Solar Neighborhood with Gaia Edr 3: Its Relation to Cluster Dynamics.” *The Astrophysical Journal*, vol. 912, no. 2, 17 May 2021, p. 162., doi:10.3847/1538-4357/abeaac.
8. Röser, Siegfried, et al. “Hyades Tidal Tails Revealed by *Gaia* DR2.” *Astronomy & Astrophysics*, vol. 621, 4 Jan. 2019, doi:10.1051/0004-6361/201834608.
9. Scudder, Jillian. “Why Doesn't a Globular Cluster Collapse?” *Astroquizzical*, Astroquizzical, 18 Aug. 2018, astroquizzical.com/astroquizzical/why-doesnt-a-globular-star-cluster-collapse-into.
10. *Star Clusters | Center for Astrophysics*. pweb.cfa.harvard.edu/research/topic/star-clusters.
11. “Why Do Clusters of Stars Eventually Dissipate?” *Physics Stack Exchange*, Stack Exchange, 27 Oct. 2011, physics.stackexchange.com/questions/25579/why-do-clusters-of-stars-eventually-dissipate.
12. “Why Do Stars Born in a Cluster Finally Disperse?” *Astronomy Stack Exchange*, 15 Sept. 2017, astronomy.stackexchange.com/questions/22707/why-do-stars-born-in-a-cluster-finally-disperse.