A Combinational Strategy for Clustering of Astronomical Datasets

[1]Hardeep Singh Kang, Research Scholar, IKGPTU, Jallandhar, India, hardeep_kang41@rediffmail.com
[2]Dr. K S Mann, Professor (Department of IT), GNDEC, Ludhiana, India, mannkulwinder@yahoo.com
[3]Mohanjit Kaur, Research Scholar, CT University, Jagraon, India, mohanjitkaur6@gmail.com

Abstract - **In this research work analysis of astronomical dataset is done for developing strategies for clustering datasets, so that more insight and information can be extracted from these datasets. The process of clustering dataset is way of organizing the datasets without the supervision of the expert or semi-automatic algorithms. Three approaches have showcased in this work using real and synthetic astrophysics data. In the first approach dynamic binning method is used to find unknown intervals problems such as red shift. The outcome clearly shows that Knuth dynamic binning is best for more information gain as compared to standardized histogram and Bayesian dynamic algorithm. The second strategy used in this context was the use of classic kmeans, heuristic kmeans and modified kmeans algorithms for astrophysics clustering problems. And the third one was use of combined algorithms to construct machine learning model (Support Vector Machine) that can classify celestial objects based on the clustered groups. The outcome of this work shows that all the three approaches are relevant and required in context of respective astrophysics problems.**

Keywords- Astronomy, Clustering, Astronomical Dataset, Unsupervised Learning

I. INTRODUCTION

Recently, research initiative by the Google and University of Texas brought dividends when they used data from NASA to identify two new exoplanets with help of machine learning (Neural Network) algorithms[1]. The algorithm produced accuracy as high as 96% inspite of the fact that the dataset has high degree of noise. The neural network scoured through the dataset that had information of 670 starts to arrive at discovery of new planets named as kepler 80 g and kepler 90i. The quest to analysis more kepler data is on and the main modus operandi is the use for unsupervised learning machine learning algorithms. In certain cases, such as pre-labeled data is available, that has tangible and easily recognizable celestial body having specific characteristic such as brightness etc. And it is beyond doubt that all these developments are happing due to progress in the image processing[2]. Therefore, any strategy in understanding the astronomical datasets can never ignore the image processing and the mathematics of searching objects in datasets[3]. Astronomy as subject cannot make progress unless it adapts and extends methods and developments from other fields[4].

The availability of the astronomical data is easy now but the initiatives to analysis such volumes of data are slow[3][5]. This may be attributed to the fact, that raw data requires lot of handling before it can give meaningful information. For doing research work in astrophysics there are two possible directions, once is Forward Modeling in Astronomy. It is an approach in which model is simulated using existing knowledge of system in terms of its environment, physics, biology mathematical models etc. or equations. This is basically a way of producing simulated data from the inputs that can be compared with the actual data. And the second is Inverse Modeling in Astronomy[6], if we have analytical data and methods that can measure certain parameters such as density of moon from the measurement of its gravity field taken using actual instruments. It is called an inverse[6] problem because it starts with result dataset and tries to identify the reason for such results or outcomes. Using these strategies the many researchers have done extensive work. The next section, the review section discusses the application of such approaches to the problems and challenges in astronomy.

II. LITERATURE REVIEW

It is no secret that the astronomical datasets are big in volume and variety. Due this intrinsic fact the data in astronomy normally need to undergo some degree of pre-processing[7] and dimension reduction[8]. Following are the methods by which meaningful information may be extracted from the dataset and these are kind of Inverse Modeling[9] in Astronomy, where we have existing store of raw data and we try to build an information model around it.

The insights provided by the dynamically programmed histograms[10][11][12] are important in astronomy, if we are analyzing a large area in the space. They can be useful in estimation of multiple parameters, especially when the data is dynamic and is changing with arbitrary or unknown interval. Choosing right interval period becomes a tradeoff

 DOI : 10.18231/2454-9150.2018.0740

ISSN : 2454-9150 Vol-04, Issue-06, Sep 2018

between getting outdated data or having unnecessary overhead. Hence, the need to divide the data using some binning algorithm is required. These algorithm divide or cluster[13] the data into separate intervals so that periodicity (e.g. periodograms[14],[15],[16][17][18][19]) in the data can more be visible or can be estimated and at the same time a particular parameter can be studied. These methods may also require fitness functions to define an optimal binning/sizing for the data For example , the study of the light curves or light intensity[20][21] is required for studying the various characteristics of the celestial bodies in space**.** These light curves[22] helps to compute the periodic variability of the body which are depicted with help period grams, which is method of clustering the data set based on periodicity.

The second strategy that can be used to get insights from the datasets using methods that finds trend lines, hyper planes and margins to divide the data for information gain. These methods are typically regression methods that are further modified to do classification job. For example , the estimation of red shifts[23] from the colors of the moving space objects such as galaxies and quasars[24] can be done using nearest neighbors [25][26][27][27][7]clustering algorithm . The clustering algorithms that compute the similarity of the objects based on distance are extremely useful in solving problems such as redshifts.

The third strategy that can used to construct engines or systems that can identify astronomical objects based on the characteristics is by using combined approach[28]. Initially, when the knowledge discovery[29],[30],[31],[32],[33] need to be done a clustering method such as K-means may be applied for find meaningful clusters. And later, label this clustered data to produce a supervised dataset that can be subjected to the machine learning algorithm such as neural networks[34].[35],[36],[37][38],[38][39] and support vector machines[34]. For example, the authors[40], did clustering process on the Sloan Digital Sky survey data based on the colors intensities that subtlety helps to distinguish dwarfs from giants even if they are have similar temperature.

Key Gaps

After conducting this review, following gaps and challenges need to be noted and addressed.

- There is a need for some degree of preprocessing in most the datasets[41],[42] of the astronomy.
- There is need for characterization of the dataset in astronomy datasets as in many cases Aprori information is not available.
- The characterization of the parameters based on which the knowledge discovery[31] can be done from the astronomical dataset need careful formulation and calibration.
- Neural network[34].[35],[36],[37][38],[38][39] may be used for identification of patterns in dataset but they may higher degree of overhead as compared to probability approaches of knowledge discovery.

III. ASTROPHYSICS PROBLEMS UNDERTAKEN

Estimating the red shifts from the colors of the moving celestial bodies is one of the primary tasks that are required by the astronomy. The computations of the red shifts can be measured with the help of photometry. In most cases the red shifts are computed by analyzing the emission spectrum at a range of red shifts but recently artificial neural networks[34],[35],[36],[37],[38],[38],[39] and probability statistics have also been used .This work focuses on the use of distance based clustering[43][44] algorithm for computing red shifts and also find optimal methods of building dynamically bins or segments for more information gain.

The brightness values[40] of celestial bodies such as planets give lot of information about space . By clustering such celestial bodies based on brightness in appropriate groups many inferences can be found .The is the second astrophysics problem, we will be undertaken.

IV. IMPLEMENTATION

This section explains the steps taken to develop clustering strategies for aforementioned problem and Astroinformatics datasets having unknown time intervals. The block diagram gives the steps in involved in conducting this work.

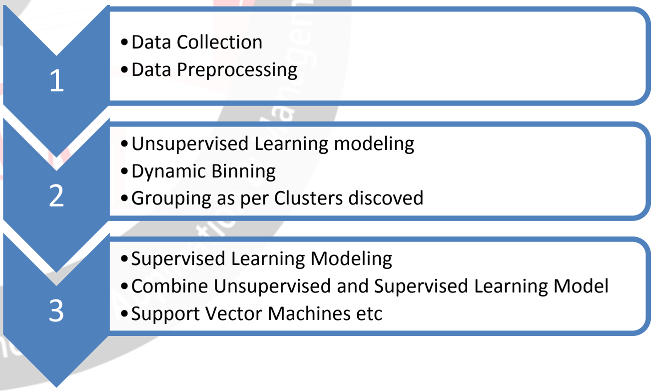

Figure 1: Flow of Work

Determining how many bins/intervals a histogram should have is an estimation non-parametric statistics problem. And by definition bin size can be defined as group, class, category or an interval to sort the data. In more general sense, histogram is a function that counts the number of observations in each disjoint category. And defining the length of this disjoint set is main challenge. However, there are mainly three possibilities ways by which the clustering astrophysics data can be done in context of our problem. These include the dynamic programming or binning methods, distance based clustering and third one is the hybrid method of clustering. The next section gives

DOI : 10.18231/2454-9150.2018.0740

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-06, Sep 2018

information on all of these types of clustering the data having unknown time interval.

1. **Dynamic Binning,** in many cases, the astrophysics data interval length is not properly known and there are fair chances that loss of information may happen while understanding the time series [45],[46],[47][48]. Hence, the need to finding appropriate method for dividing or clustering the data into intervals. In such cases the dataset is sensitive to noise also . An addition of some noise or artifact in the image single can lead to misrepresentation of the time series data sets .Previous methods have used either standardized or fixed intervals to divide the data. But the contemporary methods use fitness function(s) to divide the dataset into more meaningful bins. The Knuth method uses estimation of density of parameters in the astrophysics dataset. This algorithm is useful in cases where capturing the various density estimations is paramount and in case the dataset is small and have some degree of "Aprori" information for finding the bin size the Bayesian block method can be used. However, the generalized form of both algorithms is as follows.

Generalized Pseudo Logic (Knuth &Bayesian) :

```
bin_size = x  ;fitness_error= 0;error_tolerance = 0.2;rate_change= 0.1max_iteration_allowed=0 ;
for i= 0   to max_iteration_allowed
for i=1:r
  for j= 1:c
        Get Default Bin Size
        Segregate Data as per Default bin size ,
        Compute /Measure goodness of fitness (RMS, MISE)
        Find fitness_error
        If fitness_error< error _tolerance
          Accept new_bin_size = bin_size
        Segregate Data as per new_bin_size;
        else
           bin_size = = bin_size+rate_change
      end if
      end
      end
```

The optimal bin size is found by using iterative method .The algorithm require initial default values of bin size , allowable/tolerance error that can give optimal bin size when fitting function is executed . Other than these, it also needs the rate at which the algorithm must change (rate of change) in attempt to find optimal bin size. Mean Integrated Square Error is found to finally arrive at the decision to accept or reject the new or optimized bin size.

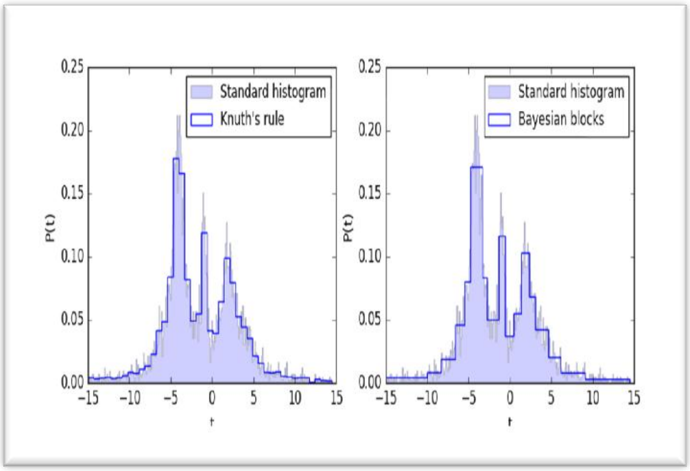

Figure 2: Dynamic Histogram method of Segmentation Output Comparison

Interpretation of the Graphs: These graph shows that the Knuth algorithm makes a narrow histograms and Bayesian algorithm makes wider interval histograms. And both these are giving better information representation with respect to the standard histograms methods .It is apparent from the above histograms that the Knuth algorithm for works well whenever , there is a need to find variability of the data. The variability in the astronomical dataset is typical characteristics. But the Bayesian or Probability approach divides the dataset better in case the analysis is quantitative in nature as it is able to represent all the features of the underlying data. This may be attributed to the fact that the standard histogram method used mean and variance statistics for segregating the datasets whereas the Knuth and Bayesian method use fitness function to make various segments of the data. Both these methods are promising method that can be utilized in astronomy in context of unsupervised machine learning approaches to handle data.

2. **Distance based Clustering:** K-means and Nearest Neighbor are most used clustering method for segmenting the dataset in astronomy. This is because the Kmeans algorithm provides invariant data and reduced size of the data similar to the PCA method besides providing the group of similar objects. In astronomy the objects in space can be groups based on size, brightness, luminance, density, frequency, spectrum characteristics and many other factors. However, all both these algorithm may require fine tuning and parameterization to be accurate enough. In certain cases there might be need for Hierarchical Clustering to arrive at conclusive decision in identification of objects in astrophysics data. The implementation of the heuristic Kmeans and modified k-means show that by modifying (reducing) the number of clusters parameter we can fine tune the distance based clustering algorithm to suit astronomy data conditions.

 DOI : 10.18231/2454-9150.2018.0740

ISSN : 2454-9150 Vol-04, Issue-06, Sep 2018

The speed of the algorithm is computed using ration of the Running Time Of Conventional Kmeans/ Running Time of Heuristic Kmeans. The dataset size was varied to evaluate the algorithm.

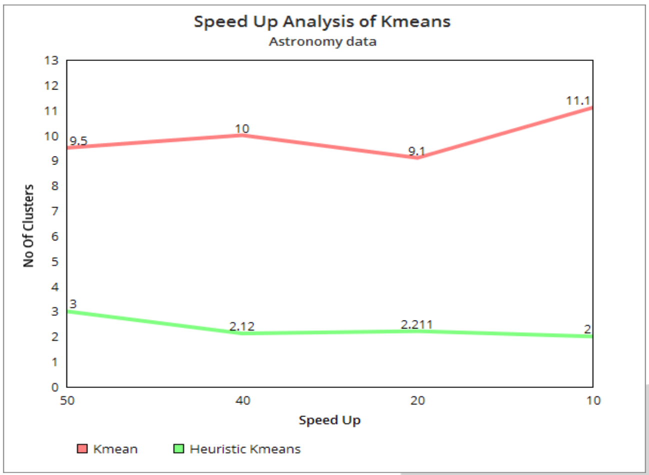

Figure 3: Speed Up Analysis of Kmean Clustering Algorithm

Interpretation of Graphs: One of the major requirements in astronomy is group similar objects having small discriminant or difference in their features. And algorithms such as K-means need to perform ‘n x k’ distance computations. If the datasets consists of n points having ‘k’clusters. To avoid such huge overhead for computing similar objects, A new combinational approach can be applied as it would help to reduce the overall overhead by decreasing the number cluster it has to iterate. The logic comes from the empirical outcomes from the various simulations that the data points have predisposition to go to collections that are closer in the preceding iteration of the distance based clustering algorithms such as Kmeans and Nearest Neighbor Algorithms. This approach shows that the accuracy of the clustering algorithm is not affected and at the same time the processing huge astrophysics dataset can be done properly

3. Strategy III

Combining Clustering and Classification: In cases, where the data has lots of insights but they need to be discovered. In cases where the expertise to group the data are not available or there the volume of dataset is huge there is a need to work with strategy III. The advantage is the discovered insights can be labeled intelligently and later this data can be used to construct supervised learning model.

Number of Clusters	Kmeans Speed Up	Heuristic Kmeans Speed Up
50	9.5	3
40	10.00	2.12
20	9.100	2.211
10	11.1	2.000

Table 1: Speed up Comparison

However it must be noted that this strategy works well, especially when the accuracy of the clustering algorithms is almost hundred percent in terms of true positives and false negatives. In such cases, the data set after the clustering can be labeled or marked with category/group/class. And a supervised machine learning algorithm may be applied to build a fully automated system that used this approach.

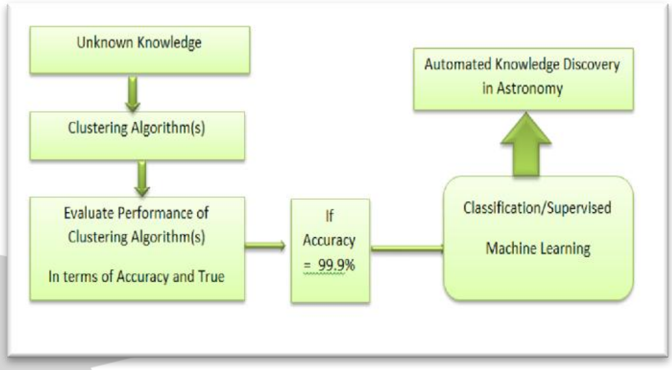

Figure 4: Combination Approach for Knowledge Discovery in Astronomy

The use of neural networks and other algorithms related to deep learning have lot of advantages, if the organization or research has lot of computation resources at his or disposal. But, for armature or for academic study the support vector machine kernel functions can come handy. As mentioned in the above section the output from the clustering algorithm can be subjected to classification algorithm for making it an automated system . The figure 5 shows the output of the SVM algorithm applied to a simulated astrophysics dataset having brightness values of celestial bodies such as planets.

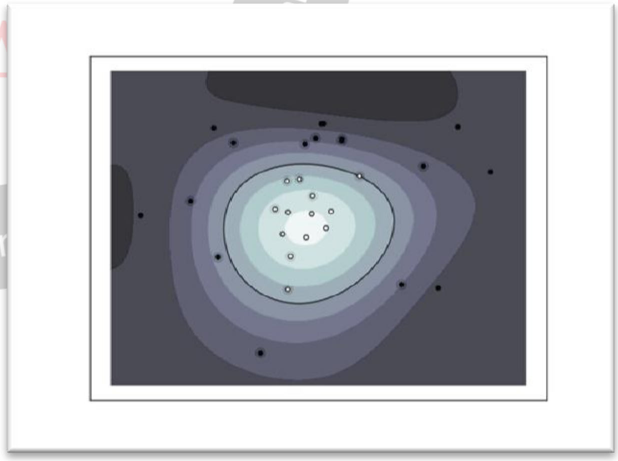

Figure 5: Support Vector Machine

Interpretation of Graphs: Based on the data sets in the astrophysics it was found that linear SVM is less disposed to over fitting than non-linear kernel methods. And typically, if we have really large compared to the training sample, linear kernel produces data learned from the grouped have wide margins and if the numbers of features are relatively less in number but the training sample is

DOI : 10.18231/2454-9150.2018.0740

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-06, Sep 2018

large, we may require to move to non-linear kernels such as radial basis or polynomial kernels.

V. RESULTS AND DISCUSSIONS

In this work a time series clustering problem is undertaken to drive insights from the data. The initial step is to find the way the data is represented. How is the time frame defined and on what logic. Do we have some prototype to represent the cluster or the group. After finding answers to all these question an intuitive clustering approach can be built. The problem however, in our context was that the time series data does not have fixed or defined time period. The time interval needs to predicted from the data pattern itself. Hence , finding patterns in the dataset or segmenting into meaningful information bytes or doing parameterized statistical clustering or even grouping the similar object to arrive at some inference is focus of this paper. One of the pressing problems in astrophysics is analysis of the brightness properties of the celestial. The problem become complex when these bodies are in motion and it is hard to define the time interval based on which some inference can be found and verified.

In this paper we are analyzing the brightness of the celestial bodies to group them using combined / hybrid method of clustering .It was also assumed that data points within a particular time frame can be treated just like an image signal. An image that has multiple bright spots having good variation. It was found in the preliminary observation that although the classical histogram analysis can help in finding good information about the data but still lot of insights and information remain hidden dues to the inadequacies of the standard method. Moreover, in standard histogram method the time interval is known and is normally constant for the said dataset. But, when the time interval is unknown Dynamic Binning as mentioned earlier in the paper performs well. The Knuth and Bayesian methods works quite well as per the outcomes of this work. Experiments with the Kmeans and Nearest Neighbor also show that these algorithms can be employed to astrophysics problems. But, the fine tuning of the k-means for astrophysics problems give better results as it is able to handle more clusters as compared to the conventional Kmeans. The last but not the least , from the interpretation of the graphs of the application of SVM on the clustered data also show that Support Vector Machine Linear function is also useful for astrophysics classification problems .

VI. CONCLUSION

From this research work it is thoroughly clear that the cluster analysis depends upon the type of Data types of the variables involved. An astronomical dataset have values that are binary, nominal, ordinal or may be ratio scaled or type data variables. Each type requires a pretreatment before it can be subjected to the cluster analysis. However, the process followed here was to find subtypes or groups that are not defined a priori based measurement. In this article, we conducted a survey of relevant possible clustering methods that can work on the astronomical datasets. The progress in this context try to seeks various combinations of measurements and errors such as positions of celestial bodies, chemical abundances magnitudes, radial velocities, stellar parameters, metallicities etc. And it was found that, different measurements require different strategies for extracting meaningful information.

VII. FUTURE DIRECTIONS

In this research work, we have done an explorative study to find which strategy works well the dataset of astrophysics problems. The approach used here can also be called “Predictive Clustering Approach” because the clustering is done to predictive various types of groups as the unknown time interval. This work can be further improved by employing such approach on other problems of astrophysics where certain pattern or insight need to first discovered and then later a supervised machine learning needs to be done. However, two things need to be taken care in future work. First, training and modeling a classifier on cluster analysis output means that by scheme or design the "groups or classes" are separable and overlapping is less. Validation need to done in a general way that it validates each steps properly.

REFERENCES

[1] Reuters, “Google crunches data to help NASA find two new planets - The Hindu,” *The hindu*, 2017. [Online]. Available: https://www.thehindu.com/sci-tech/science/nasas-kepler-probe-discovers-planet-orbiting-sun-like-star/article21666932.ece. [Accessed: 07-Sep-2018].

[2] G. B. Berriman and J. C. Good, “The application of the montage image mosaic engine to the visualization of astronomical images,” *Publ. Astron. Soc. Pacific*, vol. 129, no. 975, p. 58006, 2017.

[3] R. J. Brunner, S. G. Djorgovski, T. A. Prince, and A. S. Szalay, “MASSIVE DATASETS IN ASTRONOMY.”

[4] A. A. Goodman, “What is WorldWide Telescope, and Why Should Researchers Care?,” in *American Astronomical Society Meeting Abstracts*, 2016, vol. 227.

[5] B. collaboration and others, “The Second Data Release of the Beijing-Arizona Sky Survey,” *arXiv Prepr. arXiv1712.09165*, 2017.

[6] B. L. Ellerbroek and C. R. Vogel, “Inverse problems in astronomical adaptive optics,” *Inverse Problems*. 2009.

[7] R. Vilalta, K. D. Gupta, and A. Mahabal, “Star classification under data variability: An emerging challenge in astroinformatics,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9286, pp. 241–244.

[8] Y. Zhang and Y. Zhao, “Astronomy in the Big Data Era,” *Data Sci. J.*, vol. 14, no. 0, p. 11, 2015.

[9] A. M. Stuart, “Inverse problems: A Bayesian perspective,” *Acta Numerica*. 2010.

[10] N. Thaper, S. Guha, P. Indyk, and N. Koudas, “Dynamic multidimensional histograms,” in *Proceedings of the 2002*

 DOI : 10.18231/2454-9150.2018.0740

ACM SIGMOD international conference on Management of data - SIGMOD '02, 2002.

[11] K. M. Hammouda and M. S. Kamel, "Incremental document clustering using cluster similarity histograms," in *Proceedings - IEEE/WIC International Conference on Web Intelligence, WI 2003*, 2003.

[12] Q. Li and C. North, "Empirical comparison of dynamic query sliders and brushing histograms," in *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, 2003.

[13] W. Jang and M. Hendry, "Cluster analysis of massive datasets in astronomy," *Stat. Comput.*, 2007.

[14] G. Bretthorst, "Frequency estimation and generalized Lomb-Scargle periodograms," *Stat. Challenges Astron.*, 2003.

[15] G. B. Berriman, E. Deelman, G. Juve, M. Regelson, and P. Plavchan, "The Application of Cloud Computing to Astronomy: A Study of Cost and Performance," *Processing*, 2010.

[16] T. Reinhold, R. H. Cameron, and L. Gizon, "Astronomy Astrophysics Evidence for photometric activity cycles in 3203 Kepler stars," *A&A*, 2017.

[17] R. V. Baluev, "PlanetPack: A radial-velocity time-series analysis tool facilitating exoplanets detection, characterization, and dynamical simulations," *Astron. Comput.*, 2013.

[18] A. Schwarzenberg-Czerny and A. Schwarzenberg-Czerny, "An astronomer's guide to period searching," *... Var. Sel. Areas ...*, 2003.

[19] R. V. Baluev, "Assessing the statistical significance of periodogram peaks," *Mon. Not. R. Astron. Soc.*, 2008.

[20] V. Bharadwaj, T. H. Reiprich, G. Schellenberger, H. J. Eckmiller, R. Mittal, and H. Israel, "Intracluster medium cooling, AGN feedback, and brightest cluster galaxy properties of galaxy groups," *Astron. Astrophys.*, 2014.

[21] O. Laurino, R. D'Abrusco, G. Longo, and G. Riccio, "Astroinformatics of galaxies and quasars: A new general method for photometric redshifts estimation," *Mon. Not. R. Astron. Soc.*, vol. 418, no. 4, pp. 2165–2195, 2011.

[22] P. J. Marshall, C. J. Lintott, and L. N. Fletcher, "Ideas for Citizen Science in Astronomy," *Annu. Rev. Astron. Astrophys.*, vol. 53, no. 1, pp. 247–278, 2015.

[23] R. Schlichenmaier and M. Collados, "Astronomy & Astrophysics Spectropolarimetry in a sunspot penumbra Spatial dependence of Stokes asymmetries in Fe I 1564.8 nm," *A&A*, 2002.

[24] Á. A. Orsi, N. Fanidakis, C. G. Lacey, and C. M. Baugh, "The environments of high-redshift radio galaxies and quasars: Probes of protoclusters," *Mon. Not. R. Astron. Soc.*, 2016.

[25] R. Liu, H. Wang, and X. Yu, "Shared-nearest-neighbor-based clustering by fast search and find of density peaks," *Inf. Sci. (Ny).*, 2018.

[26] P. Indyk and R. Motwd, "Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality," *Proc. thirtieth Annu. ACM Symp. Theory Comput. ACM*, 1998.

[27] X. Zhou *et al.*, "Visualization technology and its application for massive astronomical data analyses," in *Proceedings - 8th International Conference on Intelligent Networks and Intelligent Systems, ICINIS 2015*, 2016, pp. 105–108.

[28] S. Basu, "Semi-supervised Clustering: Learning with Limited User Feedback," *PhD Texas Austin*, 2003.

[29] J. Milli *et al.*, "Discovery of a low-mass companion inside the debris ring surrounding the F5V star HD206893," *Astron. Astrophys.*, 2016.

[30] U. M. Fayyad, "Data Mining and Knowledge Applications in Astronomy Discovery in Databases : Science and Planetary," *AAAI-96 Proc.*, 1996.

[31] M. Brescia, G. Longo, and F. Pasian, "Mining knowledge in astrophysical massive data sets," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, 2010.

[32] M. Brescia, S. Cavuoti, G. S. Djorgovski, C. Donalek, G. Longo, and M. Paolillo, "Extracting knowledge from massive astronomical data sets," in *Springer Series in Astrostatistics*, 2012.

[33] H. Abbasi, M. Wolf, and K. Schwan, "LIVE data workspace: A flexible, dynamic and extensible platform for petascale applications," in *Proceedings - IEEE International Conference on Cluster Computing, ICCC*, 2007, pp. 341–348.

[34] N. M. BALL and R. J. BRUNNER, "Data Mining and Machine Learning in Astronomy," *Int. J. Mod. Phys. D*, vol. 19, no. 07, pp. 1049–1106, 2010.

[35] R. Tagliaferri *et al.*, "Neural networks in astronomy," *Neural Networks*. 2003.

[36] D. George and E. A. Huerta, "Deep neural networks to enable real-time multimessenger astrophysics," *Phys. Rev. D*, 2018.

[37] M. Hobson, P. Graff, F. Feroz, and A. Lasenby, "Machine-learning in astronomy," *Proc. Int. Astron. Union*, 2015.

[38] P. Graff, F. Feroz, M. P. Hobson, and A. Lasenby, "SKYNET: An efficient and robust neural network training tool for machine learning in astronomy," *Mon. Not. R. Astron. Soc.*, 2014.

[39] B. Naul, J. S. Bloom, F. Pérez, and S. van der Walt, "A recurrent neural network for classification of unevenly sampled variable stars," *Nat. Astron.*, 2017.

[40] I. Zehavi *et al.*, "Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data," *Astrophys. J.*, 2002.

[41] J. H. Groh, G. Meynet, and S. Ekström, "Massive star evolution: luminous blue variables as unexpected supernova progenitors," *Astron. Astrophys.*, 2013.

[42] P. Guevara *et al.*, "Robust clustering of massive tractography datasets," *Neuroimage*, 2011.

[43] L. Ertoz, M. Steinbach, and V. Kumar, "A new shared nearest neighbor clustering algorithm and its applications," *... Data its Appl. ...*, 2002.

[44] Z. Izakian, M. Saadi Mesgari, and A. Abraham, "Automated clustering of trajectory data using a particle swarm optimization," *Comput. Environ. Urban Syst.*, 2016.

[45] J. D. Scargle, "Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data," *Astrophys. J.*, 1982.

[46] J. D. Scargle, J. P. Norris, B. Jackson, and J. Chiang, "Studies in astronomical time series analysis. VI. Bayesian block representations," *Astrophys. J.*, 2013.

[47] J. D. Scargle, "Studies in astronomical time series analysis. III - Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data," *Astrophys. J.*, 1989.

[48] J. D. Scargle, "Studies in Astronomical Time Series Analysis. V. Bayesian Blocks, a New Method to Analyze Structure in Photon Counting Data," *Astrophys. J.*, 1998.

 DOI : 10.18231/2454-9150.2018.0740