Assignment 6: a mathematical essay on support vector machines (SVM)

Krishna Sah Teli
Cyber-Physical Systems (I2MP)
Indian Institute of Technology (IITM)
Madras - 600036, India
ge22m018@smail.iitm.ac.in

***Abstract*—In this paper, SVM has been studied intensively with an application exploring classification of pulsar stars based on its different properties. It explains about the different parameters of radiation emitted by pulsar stars over the period of time which play significant role in its prediction. The correlation between the prediction and other parameters have been depicted graphically with the help of data visualization tools. It shows that the parameters 'IP Skewness' and 'IP Kurtosis' are significantly and positively correlated with the target variable where as 'IP Mean' is significantly and negatively correlated. Correlation analysis also shows that all of features contribute significantly in the prediction of pulsar stars. With the help of a support vector machines, the pulsar star prediction dataset has been implemented to build a classification model. It has been demonstrated using different kernels such as linear, rbf and sigmoid and their respective accuracy, precision and recall properties have been studied.**

***Index Terms*—SVM, Skewness, Kurtosis, Correlation, Sigmoid, Rbf**

I. Introduction

The technology has been growing rapidly and its applications have great impact on our day-to-day work. It has been playing great role in modernizing every aspects of our life from industrial to personal point of view. It has been possible because of data. Data is the fuel for the today's technology. Many industries and companies came to know the importance of data during the covid-19 time as they were not having tangible access other than online to reach out to customers. By preprocessing and feature engineering techniques, raw data can be converted into meaningful data which later used for training different machine learning models or developing artificial intelligence based devices. ML and data science have geared up the revolutionary changes in the field of technology. It has not only contributed in the technical field but we can see its impact on every field from agriculture to astronomy [11], [12]. Thus, it has become the most trending topic in current time period. Ml-model can predict, estimate, recognize, detect, etc. [1], [2] based on the data and type of training model. [11]–[13] Pulsars are a rare type of Neutron star that produce radio emission detectable here on Earth. As pulsars rotate, their emission beam sweeps across the sky, and when this crosses our line of sight, produces a detectable pattern of broadband radio emission. As pulsars rotate rapidly, this pattern repeats periodically. Thus pulsar search involves looking for periodic radio signals with large radio telescopes. Each pulsar produces a slightly different emission pattern, which varies slightly with each rotation. Thus a potential signal detection known as a 'candidate', is averaged over many rotations of the pulsar, as determined by the length of an observation. In the absence of additional info, each candidate could potentially describe a real pulsar. However in practice almost all detections arecaused by radio frequency interference (RFI) and noise, making legitimate signals hard to find. In this paper, pulsar stars prediction data has been used to classify the pulsar stars based on different statistical features. It also explains the correlation between the data features such as IP Mean, IP Sd, IP skewness, IP Kurtosis, etc. with the target class. With the help of data analysis and visualization tools, data has been explored and analyzed graphically followed by development of classification model.

To build the classification model, Support Vector Machines (SVM), a supervised machine learning model [1] has been implemented. SVM is one of the powerful machine learning algorithms for classification, regression and outlier detection purposes. An SVM classifier builds a model that assigns new data points to one of the given categories. Thus, it can be viewed as a non-probabilistic binary linear classifier. It finds a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence. It has several application in industries from small scale company to big company. for example, it can be implemented for intrusion detection [5]–[7] , image processing and pattern recognition [2], medical checkup [3], [9], forecasting, developing AI-based devices, prediction in astronomy [11]–[13], scientific studies [10], etc.

The main focus in this paper is to analyze the pulsar stars prediction dataset with the help of visualization tools and data science and develop a good SVM model to classify pulsar stars. With the help of correlation analysis and techniques, the correlation between the features of stars have been illustrated graphically. It also illustrates implementation of different kernels in SVM over the data to develop a classification model. It has been found out that the accuracy of the model developed using rbf kernel with value of C 1.0 is (94.39%) and it increases as the value of c is increased and at c=1000, accuracy becomes 95.13%.

The rest of the paper has been organized as follows: section II outlines the data set related to our problem, section III explains the mathematics behind the decision tree and tools used for developing our model. section IV Modelling and section V concludes the paper with key contributions made and directions for future works.

II. Data set

The pulsar stars prediction data source is kaggle which is famous for the data science and storage of data. This data has totally 9 features namely 'IP Mean', 'IP Sd', 'IP Kurtosis', 'IP Skewness', 'DM-SNR Mean', 'DM-SNR Sd', 'DM-SNR Kurtosis', 'DM-SNR Skewness', and 'target_class'. According to the correlation analysis, all the features are significantly correlated with target variable. Hence all the features have been taken into consideration to build a model. These features show the statistical properties of radiation emitted by pulsar stars over the period of time. Based on these statistical data, star is classified into pulsar star or not a pulsar star.

From the fig.1 we can clearly see that our dataset is suffering from imbalancing as there are about 90.8% of entries are for one class of target variable and about 9.2% of entries are for other class of target variable. It can be mitigated by using sampling function of imblearn which creates the duplicate sample to balance the imbalance data.

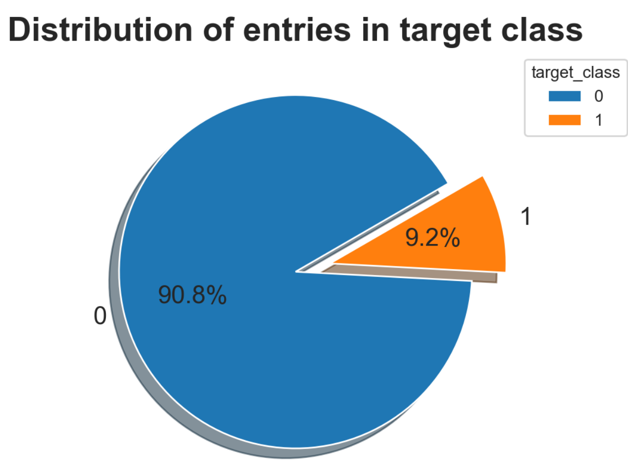

Fig. 1. Distribution of entries in target variable

With the data analysis and data visualization tools, the correlation among the data features has been studied and correlated to the each other and target class. The fig.2 shows that if the IP Mean value is less then there is higher probability of classifying the star as pulsar star and the probability decreases as the IP Mean goes beyond the 80. It is for sure that any star having IP Mean greater than 150 is not a pulsar star.

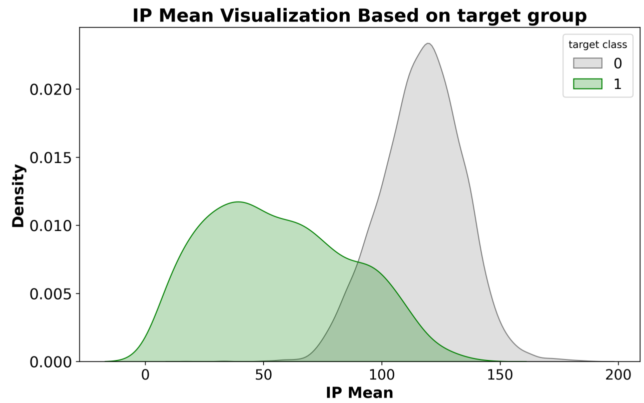

Fig. 2. Analysis of target class based on IP Mean

Similarly, fig. 3 illustrates that the a star can be categorized into pulsar star without any doubt if its IP Sd value is less than 30 and after than probability of becoming pulsar star starts decreasing. By correlating the IP Sd and IP mean, it has found that a star having less IP Mean and less IP Sd is more likely to be a pulsar star.

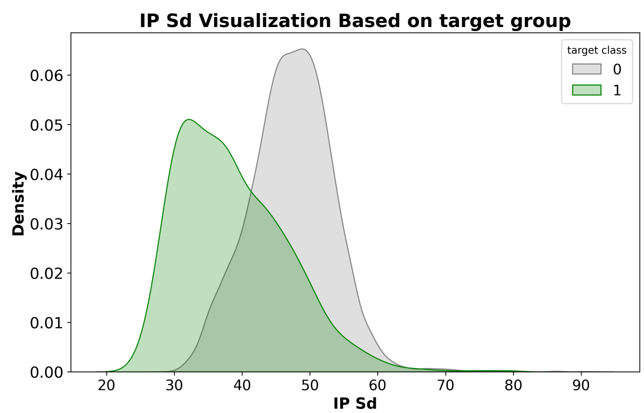

Fig. 3. Analysis of IP Sd with respect to target class

As we can see in fig.4, if the IP Kurtosis value is less i.e in range of -2 to 2 the there is higher probability of a star to be not a pulsar star otherwise it has more chance of becoming a pulsar star. However, by correlating the IP Kurtosis with IP Mean, it has found out that regardless of IP Mean value, a star having higher kurtosis i.e more than 2 then it is classified as pulsar star.

Similar behaviour is seen with the IP Skewness features with target class. From the data, it has been found that the star having DM SNR Kurtosis value less have more probability to be a pulsar star and vice versa. However, by plotting its graph again IP Mean, it has been depicted that star having IP mean less than 75 is pulsar star and otherwise not a pulsar star regardless of Kurtosis value as seen in fig.5.

With the DM SNR Mean value, it has been depicted from its visual graph that star have less probability of becoming pulsar star if it has less DM SNR Mean value and vice versa.

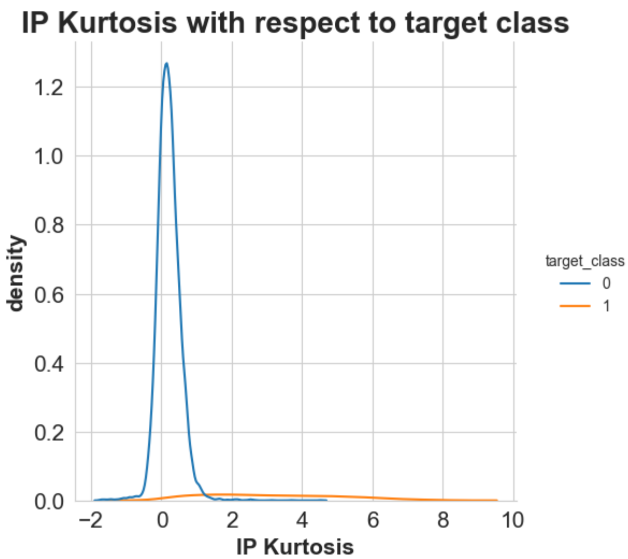

Fig. 4. IP Kurtosis against target class

Its behaviour against IP mean is similar to DM SNR Kurtosis value against IP mean.

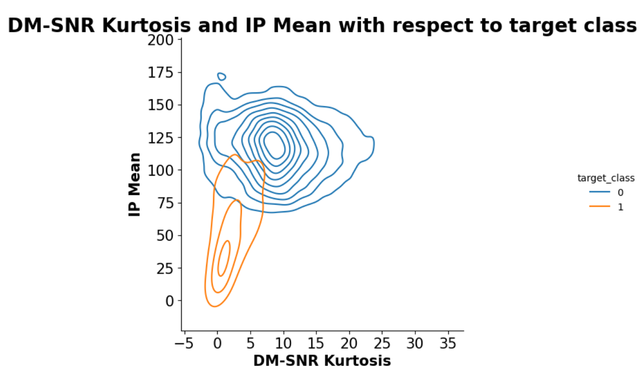

Fig. 5. DM SNR Kurtosis and IP Mean with respect to target

Similarly, DM SNR Skewness has been plotted against different other features and interesting findings we have got. As shown in the fi. 6, when DM SNR Skewness is plotted against target class depicts that the star having less value of DM SNR Skewness has more probability of classifying as pulsar star otherwise not a pulsar star. Likewise, while plotting the DM SNR Skewness against IP Kurtosis, a contradictory fact has been found out that increasing the kurtosis value and very less DM SNR Skewness value make a star pulsar star as seen in the fig.7.

When DM SNR Skewness is plotted against DM SNR Sd as shown in the fig.8, it is difficult to make classification as both plots have overlapped. It is because of outlier in the value. Similarly same output is found when it is plotted against DM SNR Mean.

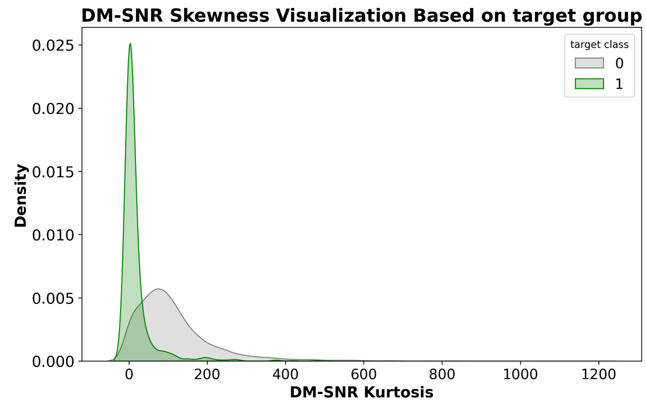

Fig. 6. DM SNR Skewness against target class

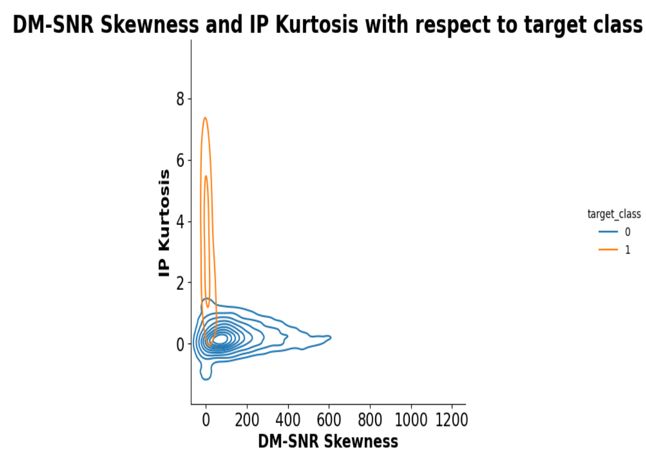

Fig. 7. DM SNR Skewness and Kurtosis against target class

III. Support Vector Machines

An support vector machine (SVM) is a supervised learning algorithm which is used for both classification and regression tasks It is one of the powerful machine learning algorithms for classification, regression and outlier detection purposes. [1]. An SVM classifier builds a model that assigns new data points to one of the given categories. Thus, it can be viewed as a non-probabilistic binary linear classifier. It finds a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. So we have to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence

[1] The original SVM algorithm was developed by Vladimir N Vapnik and Alexey Ya. Chervonenkis in 1963. At that time, the algorithm was in early stages. The only possibility is to draw hyperplanes for linear classifier. In 1992, Bernhard E. Boser, Isabelle M Guyon and Vladimir N Vapnik suggested a way to create non-linear classifiers by applying the kernel trick

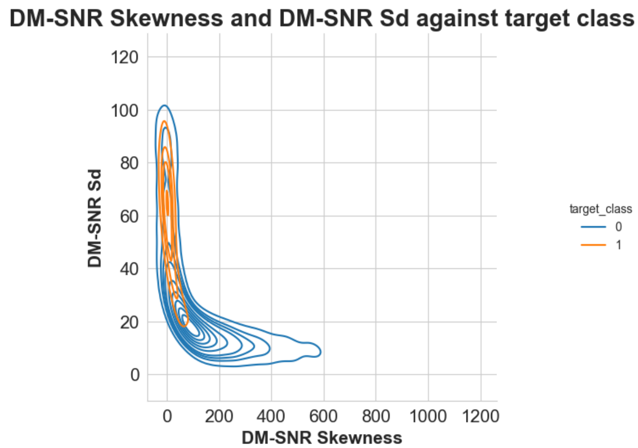

Fig. 8. DM SNR Skewness and DM SNR Sd against target class

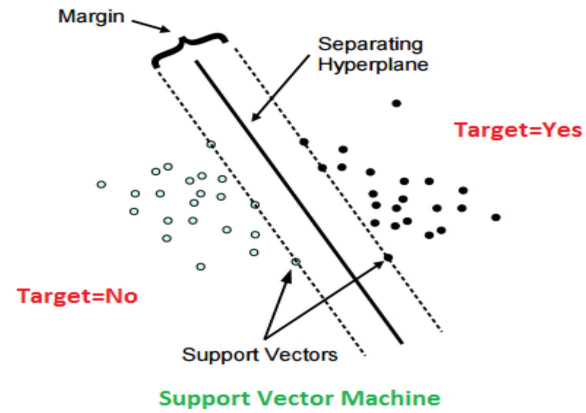

Fig. 9. Hyperplanes visualization

to maximum-margin hyperplanes. The current standard was proposed by Corinna Cortes and Vapnik in 1993 and published in 1995.

SVMs can be used for linear classification purposes. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using the kernel trick. It enable us to implicitly map the inputs into high dimensional feature spaces.

A. *Hyperplanes and Support Vectors*

Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3. The SVM classifier separates data points using a hyperplane with the maximum amount of margin. This hyperplane is known as the maximum margin hyperplane and the linear classifier it defines is known as the maximum margin classifier.

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

A margin is a separation gap between the two lines on the closest data points. It is calculated as the perpendicular distance from the line to support vectors or closest data points. In SVMs, we try to maximize this separation gap so that we get maximum margin.

SVM searches for the maximum margin hyperplane in the following 2 step process –

I. Generate hyperplanes which segregates the classes in the best possible way. There are many hyperplanes that might classify the data. We should look for the best hyperplane that represents the largest separation, or margin, between the two classes.

II. So, we choose the hyperplane so that distance from it to the support vectors on each side is maximized. If such a hyperplane exists, it is known as the maximum margin hyperplane and the linear classifier it defines is known as a maximum margin classifier.

B. *Cost Function and Gradient Update*

In the SVM algorithm, we are looking to maximize the margin between the data points and the hyperplane. The loss function that helps maximize the margin is hinge loss.

$$c(x, y, f(x)) = \left\{ \begin{array}{cc} 0 & if y * f(x) \geq 1 \\ 1 - y * f(x) & else \end{array} \right.$$

$$C(x, y, f(x)) = (1 - y * f(x))_+$$

Hinge loss function (function on left can be represented as a function on the right)

The cost is 0 if the predicted value and the actual value are of the same sign. If they are not, we then calculate the loss value. We also add a regularization parameter the cost function. The objective of the regularization parameter is to balance the margin maximization and loss. After adding the regularization parameter, the cost functions looks as below.

$$min_w \lambda ||w||^2 + \sum_{i=1}^{n} (1 - y_i < x_i, w >)$$

loss function for SVM

Now that we have the loss function, we take partial derivatives with respect to the weights to find the gradients. Using the gradients, we can update our weights.

$$\frac{\delta}{\delta w_k} \lambda ||w||^2 = 2\lambda w_k$$

$$\frac{\delta}{\delta w_k}(1 - y_i < x_i, w >)_+ = \left\{ \begin{array}{cc} 0 & if y_i < x_i, w > \geq 1 \\ -y_i x_{i*k}, & else \end{array} \right.$$

Gradients

When there is no misclassification, i.e our model correctly predicts the class of our data point, we only have to update the gradient from the regularization parameter.

$$w = w - \alpha * (2\lambda w)$$

Gradient Update- Misclassification

When there is a misclassification, i.e our model make a mistake on the prediction of the class of our data point, we include the loss along with the regularization parameter to perform gradient update.

$$w = w + \alpha * (y_i * x_i - 2\lambda w)$$

Gradient Update- Misclassification

IV. Modelling

Support Vector Machine is one of the powerful machine learning algorithms for classification, regression and outlier detection purposes [1]. The SVM classifier separates data points using a hyperplane with the maximum amount of margin. This hyperplane is known as the maximum margin hyperplane and the linear classifier it defines is known as the maximum margin classifier. Here SVM model with different kernel and different value of C have been trained and their accuracy have been analysed. With linear kernel, the accuracy of the model with C=1.0 is 93.74% and with C= 100 is 93.65%. Similarly, with sigmoid kernel, the accuracy of model with C= 1.0 is 93.74% and with C= 100 is 93.65% as of linear kernel. However, with rbf kernel, the accuracy of model increases as the value of C is increased. When C= 1.0, the accuracy is 94.39%, with C= 100, the accuracy becomes 94.78%. Finally with C=1000, the accuracy of model is found to be 95.13% and after that accuracy starts decreasing if C is increased further more. Thus, best accuracy is achieved when C= 1000 and kernel is rbf.

Acknowledgement

It is my pleasure to express with deep sense of gratitude to Professor Gaurav Raina, for his guidance, continual encouragement, understanding.

I would like to extend my gratitude to IIT Madras, for providing with an environment to work in and for his inspiration during the tenure of the course.

Conclusion

In this paper, SVM has been used to train a classification model to classify star into pulsar star or not. Data analysis and visualization has been implemented visualize the correlation among the variables and cause-effect. The SVM model was trained using different kernels such as linear, sigmoid and rbf with different value of C to find possible highest accuracy. The highest accuracy of the model has been found 95.13% with rbf kernel and C=1000. This paper also helps new researcher to get more information about SVM and related terminologies.

References

[1] N. Guenther and M. Schonlau, "Support vector machines," Stata Journal, vol. 16, no. 4, 2016, doi: 10.1177/1536867x1601600407.

[2] S. Huang, C. A. I. Nianguang, P. Penzuti Pacheco, S. Narandes, Y. Wang, and X. U. Wayne, "Applications of support vector machine (SVM) learning in cancer genomics," Cancer Genomics and Proteomics, vol. 15, no. 1. 2018. doi: 10.21873/cgp.20063.

[3] NC. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Min Knowl Discov, vol. 2, no. 2, 1998, doi: 10.1023/A:1009715923555.

[4] J. Luo, X. Yan, and Y. Tian, "Unsupervised quadratic surface support vector machine with application to credit risk assessment," Eur J Oper Res, vol. 280, no. 3, 2020, doi: 10.1016/j.ejor.2019.08.010.

[5] A. Mishra and Fancy, "Efficient Detection of Phising Hyperlinks using Machine Learning," International Journal on Cybernetics Informatics, vol. 10, no. 2, 2021, doi: 10.5121/ijci.2021.100204.

[6] A. Bilen and A. B. Özer, "Cyber-attack method and perpetrator prediction using machine learning algorithms," PeerJ Comput Sci, vol. 7, 2021, doi: 10.7717/PEERJ-CS.475.

[7] A. Yeboah-Ofori et al., "Cyber Threat Predictive Analytics for Improving Cyber Supply Chain Security," IEEE Access, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3087109.

[8] A. Kajal and S. K. Nandal, "A hybrid approach for cyber security: Improved intrusion detection system using ann-svm," Indian Journal of Computer Science and Engineering, vol. 11, no. 4, 2020, doi: 10.21817/indjcse/2020/v11i4/201104300.

[9] A. Ed-daoudy and K. Maalmi, "Breast cancer classification with reduced feature set using association rules and support vector machine," Network Modeling Analysis in Health Informatics and Bioinformatics, vol. 9, no. 1, 2020, doi: 10.1007/s13721-020-00237-8.

[10] B. Gaye, D. Zhang, and A. Wulamu, "Improvement of Support Vector Machine Algorithm in Big Data Background," Math Probl Eng, vol. 2021, 2021, doi: 10.1155/2021/5594899.

[11] F. Gieseke, "From Supervised to Unsupervised Support Vector Machines and Applications in Astronomy," KI - Kunstliche Intelligenz, vol. 27, no. 3, 2013, doi: 10.1007/s13218-013-0248-1.

[12] M. Pasquato and C. Chung, "Merged or monolithic? Using machine-learning to reconstruct the dynamical history of simulated star clusters," Astron Astrophys, vol. 589, 2016, doi: 10.1051/0004-6361/201425181.

[13] M. Murali, H. Sharma, and A. Mishra, "Pulsar Star Detection Using Ensemble Algorithms and Stack Generalization," in Lecture Notes in Networks and Systems, 2021, vol. 177 LNNS. doi: 10.1007/978-981-33-4501-0_15.

[14] T. Syriopoulos, M. Tsatsaronis, and I. Karamanos, "Support Vector Machine Algorithms: An Application to Ship Price Forecasting," Comput Econ, vol. 57, no. 1, 2021, doi: 10.1007/s10614-020-10032-2.