



# Supernova Type Ia Diversity: A Study using DBSCAN Algorithm

Neha Malik<sup>1</sup>, Vivek Jaglan<sup>2</sup>, Meenu Vijarana<sup>3</sup>, Nidhi Arora<sup>4</sup>, Shashikant Gupta<sup>5</sup>

<sup>1</sup>Amity University, Gurgaon, Haryana, India, nehag78@gmail.com

<sup>2</sup>Graphic Era Hill University, Dehradun, Uttarakhand, India, jaglanvivek@gmail.com

<sup>3</sup>Amity University, Gurgaon, Haryana, India, mvijarana@ggn.amity.edu

<sup>4</sup>BITS Pilani, India, nidhi.arora1@wilp.bits-pilani.ac.in

<sup>5</sup>G D Goenka University Gurgaon, Haryana, India, shashikantgupta.astro@gmail.com

## ABSTRACT

Type Ia Supernovae (SNe Ia) form a distinct class of objects to serve as standard candle for cosmological measurements. However, recently it has been found that the all the SNe Ia are not uniform and there could be sub-classes within the SNe Ia category. We investigate the possibility of sub-categorization of SNe Ia using Density Based Clustering Algorithm and observed the dependence of SN brightness on morphology of host galaxy. Our analysis, based on 40 SNe Ia, shows the presence of different subgroups within the SNe Ia class.

**Key words:** supernova, clustering, morphology, luminosity

## 1. INTRODUCTION

Explosion of a star is considered as gigantic event of this universe. This explosion when on its peak, outshine the brightness of whole galaxy. These explosions are termed as Supernovae (SNe) explosions. A star, during its lifetime builds number of elements mentioned in periodic table due to nuclear fusion. These elements are released in space during Supernova explosions. Thus, SNe are the source of life in this universe creating solar system, galaxies from the matter expelled into space in the form of particles into clouds of gas and dust.

Supernovae (SNe) explosion are classified as thermonuclear and core collapse depending upon type of explosion. Thermonuclear explosion results into Type Ia Supernovae whereas core collapse of a big star results into Type II, Type Ib and Type Ic Supernovae. All stars irrespective of their mass, follows same process of burning at initial stage where hydrogen is converted into helium. When hydrogen in the core gets exhausted, nuclear fusions stops, and core starts contracting. Due to the contraction, temperature of the core rise and it results in fusion of helium into carbon. When the helium also gets consumed, core again starts contracting. Next level is fusion of carbon but in low mass star, fusion process stops, resulting into remnant of degenerate carbon-oxygen core referred as white dwarf.

Type Ia Supernovae (SNe Ia) are ultimate explosion of white dwarf that accretes mass from its companion star in binary

system. After accreting when the mass of white dwarf exceeds over the Chandrasekhar limit of 1.4 solar masses, it results into Supernova Type Ia explosion. SNe Ia are considered as distance indicators due to their standardized luminosity. The inter-relation between standardized luminosity and light-curve serve a key role to understand the features of SNe Ia.

To determine classes within spatial data related to Supernovae, some competent algorithm was required. Density based spatial clustering of applications with noise (DBSCAN) has been used for the purpose as it is an unsupervised method with least dependency on domain knowledge. Moreover, with its unique feature DBSCAN can find outliers in the data also.

## 2. DATA SET

The data used in our analysis has been collated from various sources [1, 2, 4]. The final data set is presented in table 1. It contains 40 type Ia SNe along with their tag in column 1. The photometric information, the b-band absolute magnitude ( $M_B$ ) and the decline rate ( $\Delta m_{15}$ ) are presented in column 2 and 3 respectively. Information of the host galaxy, the name, morphology type and the corresponding numerical codes ( $T$ ) are available in column 4, 5 and 6 respectively.

## 2. METHODOLOGY: DENSITY BASED CLUSTERING

To determine the presence of groups in the data one needs to apply clustering algorithms [8, 9]. Density based spatial clustering of applications with noise (DBSCAN) is a widely used powerful clustering technique [6]. It is a density based non-parametric clustering algorithm which has several advantages over other clustering methods such as K-Means. Given a set of points in some space it can group the tightly packed points together by calculating the distance between the points.

Usually the Euclidean distance is considered, however, some other means of distance calculation such as Manhattan distance can also be used [7]. A brief account of the technical details of DBSCAN is given in section 3.1.

**Table 1:** Table of 40 Supernovae

<b>SN Name</b>	$\Delta m_{15}$	$M_B$	<b>Galaxy</b>	<b>Morphology Type</b>	$T$
SN1992A	1.47	-18.81	NGC1380	S0	-1.9
SN1989B	1.34	-18.87	NGC3627	SBb	3
SN2003kf	1.01	-19.37	M-02-16-02	Sb?	3
SN1996X	1.25	-19.24	NGC5061	E0	-5
SN1999ee	0.94	-19.46	IC5179	Sbc	4
SN1990N	1.08	-19.23	NGC4639	SBbc	3.8
SN1994D	1.32	-19.06	NGC4526	S0	-2
SN2003du	1.06	-18.93	UGC 9391	SBdm	8
SN2001el	1.15	-18.71	NGC1448	Sc	5.9
SN1997br	1.04	-19.62	E576-G40	SBd: pec	7
SN1999cw	0.94	-19.24	M-01-02-01	SBab pec:	1.5
SN1991T	0.95	-19.62	NGC4527	SBbc	3.8
SN1998bu	1.04	-19.12	NGC3368	Sab	2
SN1983G	1.37	-18.62	NGC4753	S0	-2.2
SN2002bo	1.17	-19.42	NGC3190	Sa pec	1
SN2002er	1.33	-19.45	UGC10743	Sa?	1
SN1984A	1.21	-19.46	NGC4419	SBa	1
SN1989A	1.06	-19.21	NGC3687	SBbc	4.1
SN2002dj	1.12	-19.05	NGC5018	E3:	-5
SN1981B	1.11	-19.21	NGC4536	SBbc	4.5
SN1999by	1.87	-16.64	NGC2841	Sb	3
SN1991bg	1.93	-16.81	NGC4374	E	-4.7
SN1997cn	1.86	-16.95	NGC5490	E	-5
SN1993H	1.70	-18.20	E445-G66	SBab	1.9
SN1986G	1.78	-17.48	NGC5128	S0	-2.2
SN1990O	0.96	-19.40	M+03-44-03	SBa	1
SN1990T	1.15	-19.17	PGC0063925	S0	-2
SN1991S	1.04	-19.24	UGC 5691	Sab	1.8
SN1991U	1.06	-19.49	IC4232	Sbc	3.8
SN1991ag	0.87	-19.40	IC4919	SBd	7.9
SN1992K	1.93	-17.72	E269-G57	SBab	1.9
SN1992P	0.87	-19.34	IC3690	Sbc	4
SN1992al	1.11	-19.47	E234-G69	SBc:	5.1
SN1992bc	0.87	-19.64	E300-G09	Sc	5
SN1992bk	1.57	-19.03	E156-G08	E	-5
SN1992bl	1.51	-19.13	E291-G11	SBa	1
SN1992bo	1.69	-18.76	E352-G57	S0/a	-1.5
SN1993ah	1.30	-19.28	E471-G27	S0	-2
SN1937C	0.87	-19.56	IC4182	Sm	8.9
SN1972E	0.87	-19.69	NGC5253	Sd	8

### 3.1 The DBSCAN Algorithm

The density-based clustering works on a set of points in given space, (defined by the data table) partition into several dense regions called clusters. Different clusters are separated by sparse regions. The algorithm requires mainly two parameters; one of them is the minimum number of points (*minPts*) required to form a cluster. The second parameter is the distance used to explore the neighbourhood of a point and is denoted by  $\epsilon$ .

All the points in the given space are visited in a specific manner. We start with an arbitrary point and explore its  $\epsilon$ -neighbourhood. If the point has enough number of other points (*minPts*) in the  $\epsilon$ -neighbourhood then that point is termed as core point, and a cluster starts forming, otherwise the point is termed as noise. All points that are within the boundary of the  $\epsilon$ -neighbourhood of that core point are termed as boundary point, and these boundary points are also included within the same cluster along with all other core points with in  $\epsilon$ -neighbourhood. Core points and boundary points required to include in the cluster are determined using the concept of direct density reachable and density reachable points. The above process is repeated until all the density connected points are found and included in one cluster. It should be noted that a point once labelled as noise while forming one cluster may be found in another cluster later so every un-visited point needs to be dealt with above mentioned process.

To determine the  $\epsilon$ -neighbourhood of a point one needs to calculate the distance between the core point and its neighbours. Usually the Euclidean distance is taken into consideration, calculated as follows in 3-dimensional Cartesian space:

$$d_E = \sqrt{x^2 + y^2 + z^2}. \quad (1)$$

In general, the space and its dimensions are determined by the data columns. Dimensions are equal to the number of columns used from the data sample and it also represents the required parameters. It is not necessary to use the Euclidean distance formula, other distance functions like Manhattan function can also be used. We have used Euclidean distance to find distance in our analysis.

One must be careful while choosing the numerical value of the parameters  $\epsilon$  and the *minPts*. For instance, if  $\epsilon$  is too small and *minPts* is large, the  $\epsilon$ -neighbourhood may not have sufficient points to fulfil the *minPts* criteria for a cluster. In such a situation most of the points may be labelled as noise. On the other hand, if  $\epsilon$  is too large then most of data points may fall in the same cluster. Both the situations do not represent the actual subgroups in the data. We have taken *minPts*  $\geq 3$  in our analysis. A technique to test the validity of DBSCAN clustering discussed below could be helpful in deciding the value of  $\epsilon$ .

### 3.2 Silhouette Score and Validity of Clustering

After the clusters have been formed one should check the consistency of these clusters. Silhouette score [5] provides a

way to test the consistency of the clusters in the data. This method can be applied on any method such as K-means or DBSCAN. It compares the similarity of a point to its own cluster known as cohesion to its dissimilarity with the nearest cluster known as separation. The cohesion for an object  $i$  in each cluster  $X$  is calculated as

$$a_i = \frac{1}{N(X) - 1} \sum_{j \neq i} d(i, j) \quad (2)$$

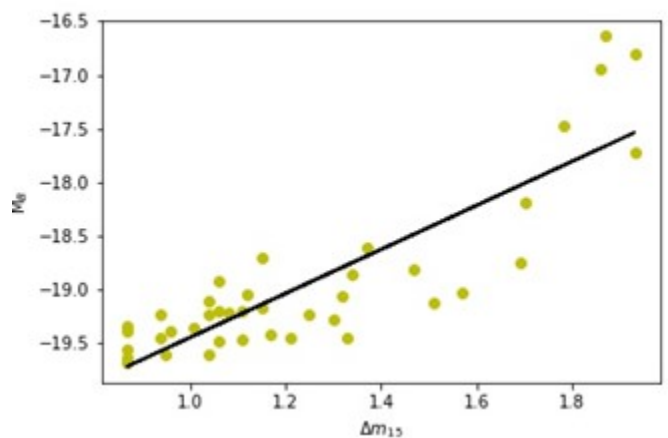
where  $N(X)$  is the number of objects in the cluster  $X$  and  $d(i, j)$  is the distance to  $j^{th}$  object of cluster  $X$ . The separation of  $i^{th}$  point to its nearest cluster  $Y$  is calculated as

$$b_i = \frac{1}{N(Y)} \sum_{j \in Y} d(i, j) \quad (3)$$

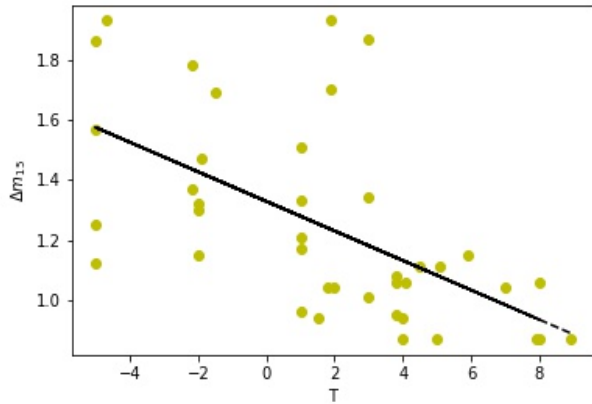
Now the Silhouette score is calculated as

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (4)$$

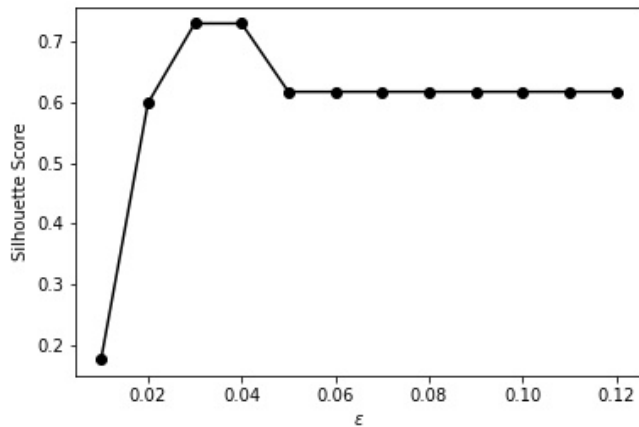
A small value of  $a_i$  and a large value of  $b_i$  shows that cluster  $X$  is cohesive and is well separated from other clusters. A positive Silhouette score is expected in this case. On the other hand,  $b_i > a_i$  leads to negative  $S_i$  and indicates that the clusters are not well separated. A small deviation in  $\epsilon$  or *minPts* will change the configuration of the clusters drastically in this case. Since, the score in Eq. 4 is normalised, its value lies between  $-1$  and  $+1$ ; and a value close to  $+1$  is favoured. By calculating the Silhouette score for various values of  $\epsilon$  one can determine a rational value of the parameter  $\epsilon$  for which Silhouette score is maximum. This value can now be used in the DBSCAN algorithm to look for subgroups in the data.



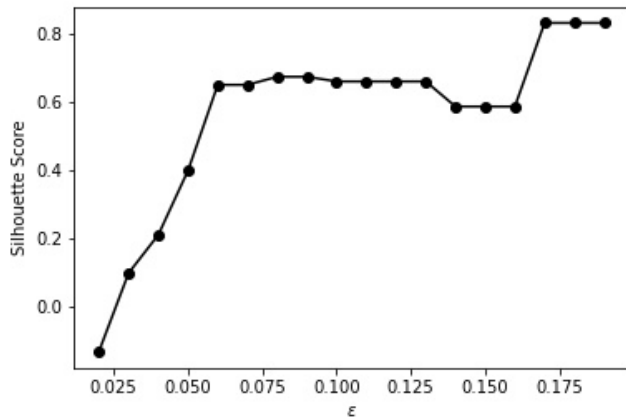
**Figure 1:** Graph shows correlation between  $\Delta m_{15}$  and  $M_B$  of 40 SNe Ia with best fit line.



**Figure 2:** Graph shows correlation between  $T$  and  $\Delta m_{15}$ .



**Figure 3:** Graphs between Silhouette-Score and Epsilon for DBSCAN algorithm on  $T$  and  $M_B$ .



**Figure 4:** Graphs between Silhouette-Score and Epsilon for DBSCAN algorithm on  $T$  and  $\Delta m_{15}$ .

**Table 2:** Correlation among various columns of data sample.

Columns	Correlation Coefficient
$\Delta m_{15} - M_B$	0.86
$T - \Delta m_{15}$	-0.59

**Table 3:** Correlations within clusters when DBSCAN is implemented on  $T$  and  $M_B$ .

Cluster no.	Data points	$\Delta m_{15}, M_B$	$T, M_B$
1	7	0.802	-0.343
2	12	0.949	-0.514
3	5	0.897	0.642
4	4	0.934	-0.317
5	10	0.837	0.664

**Table 4:** Correlations within clusters when DBSCAN is implemented on  $T$  and  $\Delta m_{15}$ .

Cluster no.	Data points	$\Delta m_{15}, M_B$	$T, M_B$	$T, \Delta m_{15}$
1	28	0.842	-0.219	-0.447
2	12	0.865	-0.288	-0.155

#### 4. RESULTS

First, we determine the correlation among various columns of the data set presented in table 1. The results have been shown with best fit line in Fig 1 and table 2. As expected, Fig 1 and 1<sup>st</sup> row of table 2 show a strong correlation between the absolute magnitude  $M_B$  and the decline rate  $\Delta m_{15}$ . This correlation indicates that bright SNe decline slowly and forms the basis of calibration of SNe Ia as standard candles for cosmological applications [3]. A slightly poor correlation is found between  $T$  and  $\Delta m_{15}$ , as seen in Fig 2 and the second row of table 2. Since,  $\Delta m_{15}$  is correlated with  $M_B$ , it shows that late type galaxies (higher values of  $T$ ) favor brighter SNe Ia.

Before implementing the DBSCAN algorithm on the data set, we need to obtain an appropriate value of  $\epsilon$ . A graph between Silhouette-Score and  $\epsilon$  has been plotted in Figure 3 and Figure 4. As discussed in section 3.2, the value of  $\epsilon$  with highest silhouette score is favored. DBSCAN algorithm is implemented on sets of two parameters using best epsilon value obtained from Silhouette Score Coefficient technique.

Now we implement the DBSCAN algorithm on the data sample of 40 Supernovae given in table 1. The value of  $\epsilon$  as shown in Fig 3 and Fig. 4 corresponds to the highest Silhouette score. When DBSCAN is implemented on the parameters  $T$  vs.  $M_B$ , it results five clusters, as shown in Fig 5. As referred in table 3, a strong correlation exists between  $M_B$  and  $\Delta m_{15}$  within each cluster. Apart from this, a relatively poor correlation is also found between  $T$  vs.  $M_B$  within each cluster.

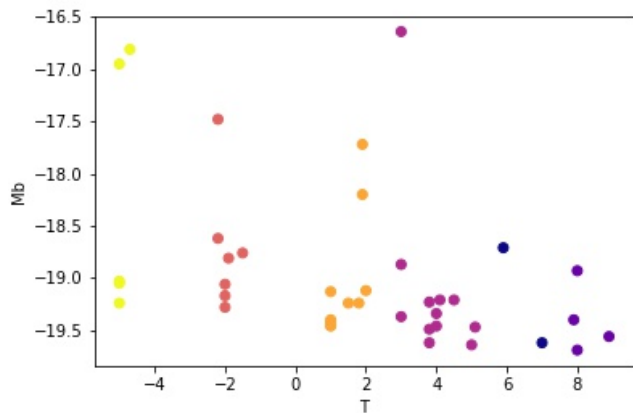


Figure 5: Graph between T and  $M_B$  with five clusters.

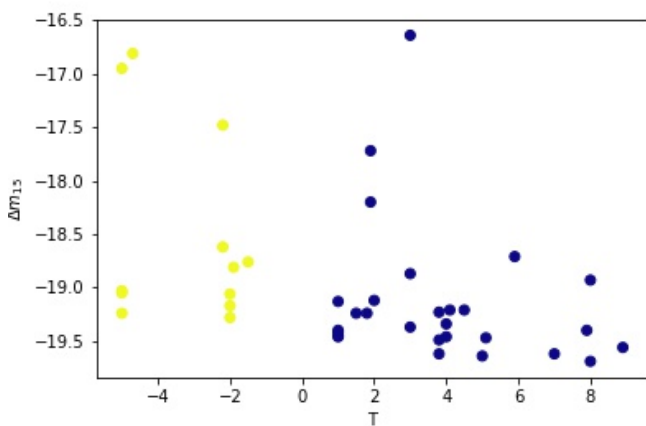


Figure 6: Graph between T and  $\Delta m_{15}$  with two clusters.

DBSCAN is also implemented on the parameters  $T$  vs.  $\Delta m_{15}$ . Now two clusters have been identified as shown in Fig 6. Table 4 specify a strong correlation between  $M_B$  and  $\Delta m_{15}$  within each cluster and relatively poor correlation between  $T$  vs.  $\Delta m_{15}$

#### 4. CONCLUSION

We have implemented the Density based scan algorithm on the SNe Ia data presented in table 1. Silhouette score is used for confirming the validity of the algorithm. For this work, we conclude that although SNe Ia form a distinct class of objects, there is a scope of sub-classes within this category. Based on the parameters selected for the analysis there exist some clusters or groups of SNe Ia. In some of these groups the correlations between the usual parameters such as galaxy morphology ( $T$ ) and the absolute magnitude ( $M_B$ ) of the SN are quite different from other groups for SNe Ia category. For instance, the overall correlation between  $T$  and  $M_B$  of SNe Ia is negative, which indicates that higher  $T$  (late type galaxies) favor lower  $\Delta m_{15}$  (fainter SNe). However, two clusters in table 3 indicates that SNe of these groups obey a reverse trend. A detailed analysis with a bigger set of SNe Ia and larger set of parameters may indicate some surprising results.

#### REFERENCES

1. Mario Hamuy et al.. **The absolute luminosities of the calan/tololo Type Ia Supernovae**, *The Astrophysical Journal*, 112, no. 6, pp. 2391–2396, 1996. <https://doi.org/10.1086/118190>
2. S.Benetti et al.. **The diversity of Type Ia Supernovae: evidence for systematics?**, *The Astrophysical Journal*, 623, no. 2, pp. 1011-1016, 2005. <https://doi.org/10.1086/428608>
3. M.M.Phillips et al.. **The Absolute Magnitudes of Type Ia Supernovae**, *The Astrophysical Journal*, 413, no. 2, pp. L105–L108, 1993. <https://doi.org/10.1086/186970>
4. Barbon et al.. **Asiago Supernova Catalogue**, *Astronomy and Astrophysics*, Supplement 81, no. 3/DEC, pp. 421, 1989.
5. Peter J.Rousseeuw. **Silhouettes: a graphical aid to the interpretation and validation of cluster analysis**, *Journal of Computational and Applied Mathematics*, 20, pp. 53-65, 1987. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
6. M. Ester, H. Kriegel, J. Sander, X. Xu. **A density-based algorithm for discovering clusters in large spatial databases with noise**, *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD'96)*, pp. 226-231, 1996.
7. Charu C.Aggarwal. **Data Mining**, Springer, pp. 153-204, 2015. [https://doi.org/10.1007/978-3-319-14142-8\\_6](https://doi.org/10.1007/978-3-319-14142-8_6)
8. S.Neware. **Fruit Grading System using k means clustering and Artificial Neural Network**, *IJATCSE*, Volume 9, No.1, January-February 2020. <https://doi.org/10.30534/ijatcse/2020/95912020>
9. Radha Mothukuri et al.. **Data Mining on Pridiction of Crime and Legal Judgements: A State of Art**, *IJATCSE* Volume 8, No. 6, pp. 3670-3679, November-December 2019.