Comparison of Classifiers to Identify the Star Type and finding Association rules using Astronomical Data

Alok Kumar*, B Shruti Mudaliar

JK Lakshmipat University

Mahapura, Ajmer Road, Jaipur 302 026, India

Abstract

Stars are celestial bodies which have played a major role in discovering theories and secrets of origin of universe and galaxies. Even today, Stars are a central subject of research for astronomers and scientists. Over the years, a lot of classification theories and models such as Morgan-Keenen spectral classification, Yerkes Luminosity classes, Hertzsprung-Russel diagram, etc have been devised which gives us star types. This paper deals with a Machine Learning approach. Different classification models have been created to identify star type (Yerkes luminosity classes). Performance of classifiers have been assessed to find out the best classifier. Further, Association rule mining has been done to get some more insights on data. Rules with high confidence and lift have been filtered for each star type.

Keywords: Association rule mining, Astronomy, Classifier, Transactional data

1. Introduction

Stars are the most well known astronomical objects. These are illuminating balls of gases which acted as a navigator for ancient discoverers. In the present scenario, these help the scientists and astrophysicists to unravel the mysteries

*Alok Kumar
**B Shruti Mudaliar
Email addresses: alokkumar@jklu.edu.in (Alok Kumar), bshrutimudaliar@jklu.edu.in (B Shruti Mudaliar)

of universe. Stars are the fundamental units of galaxies. The age, distribution and constitution of the stars uncover the history, dynamics and evolution of galaxies. Moreover, Stars are accountable for the production of heavy elements such as Oxygen, Nitrogen, Carbon, etc and the attributes are related to the attributes of the planetary systems that may converge about them. Therefore, Research and study on the birth, life and death is a focal point in the branch of astronomy. This paper is also based on stars, more precisely identification of Star type. However, the perspective presented is different.

2. Related Work

A lot of research has been done in the field of astrophysics and stars. Some of them include study of galaxies, star-forming galaxies, pulsars, quasars, asteroids, different types of stars like quark stars, neutron stars, the first stars, etc.

One paper deals with spatial clustering of star-forming galaxies in which the authors have analyzed spatial distribution of selected 28,500 galaxies and found results about the redshift and spectrum of galaxies. Another paper reviews the theories of formation of fist stars, done numerical simulations and tried to understand the accretion of these stars.

Some papers have studied clustering of stars. One of such methods is extraction of light curve sub sequences which are represented as features(vectors). Some other approaches use nearest neighbor algorithm and MST (Minimum Spanning Tree) method, one of which has been done on the star region in the small megallanic cloud.

One paper has devised a computer system to mimic MK spectral classification and the program can detect some sommon spectral-type pecularities. It can also identify WD spectra and carbon stars. This paper aims at making their system better by using new spectrograph techniques.

3. Methodology

3.1. Objective

Compare the performance of different classifiers to identify star type.

3.2. Data Understanding

The data is of 240 stars with their 7 attributes. The dataset has been taken from Kaggle. The dataset is in CSV (Comma Separated Values) and contains information about 240 stars. The author of dataset has collected the data of stars from the web. The author has found the missing values using Stefan's-Boltzmann's law to calculate the luminosity of star, Wein's displacement law to calculate the temperature and parallax to calculate radius. The data consists of 4 columns with continuous data and 3 columns of categorical data.

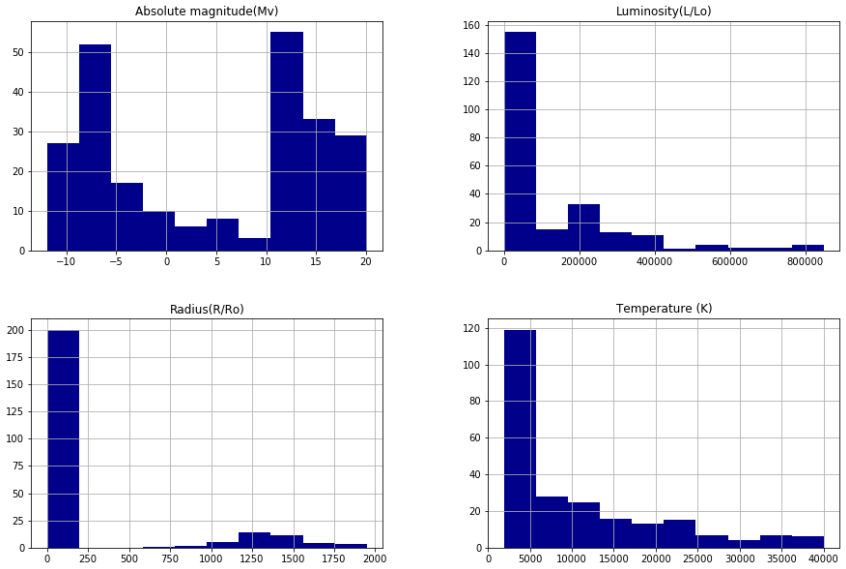
3.3. Data Preprocessing

The columns with continuous data are taken as features and Star type is the target for classification models. The association rule mining has done by dropping the inessential features and converting the data into transactional data. Conditional expressions have been used to convert the data into transactional data. Every category has been created as an attribute and each record is populated with the attribute value if it exists for the particular record. The star type (target) has the following classes:

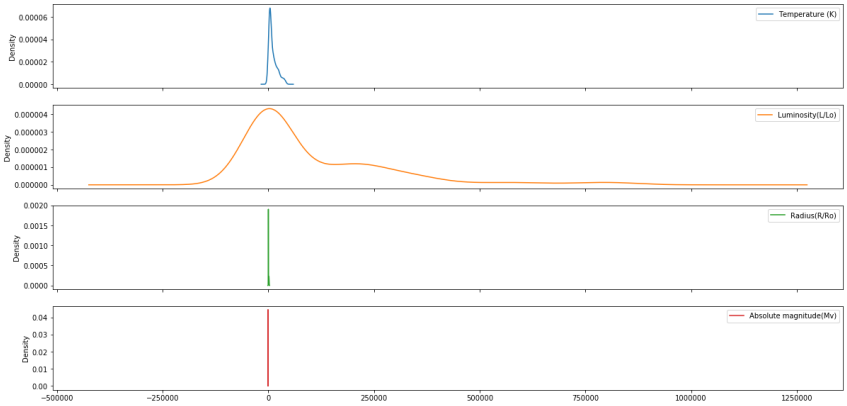
- Brown dwarf - 0
- Red dwarf - 1
- White dwarf - 2
- Main sequence - 3
- Supergiant - 4
- Hypergiant – 5

3.4. Descriptive Data Analysis

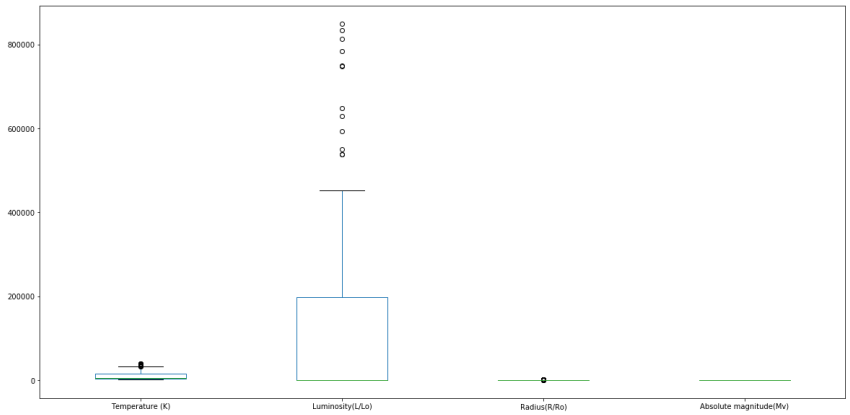
- Histogram



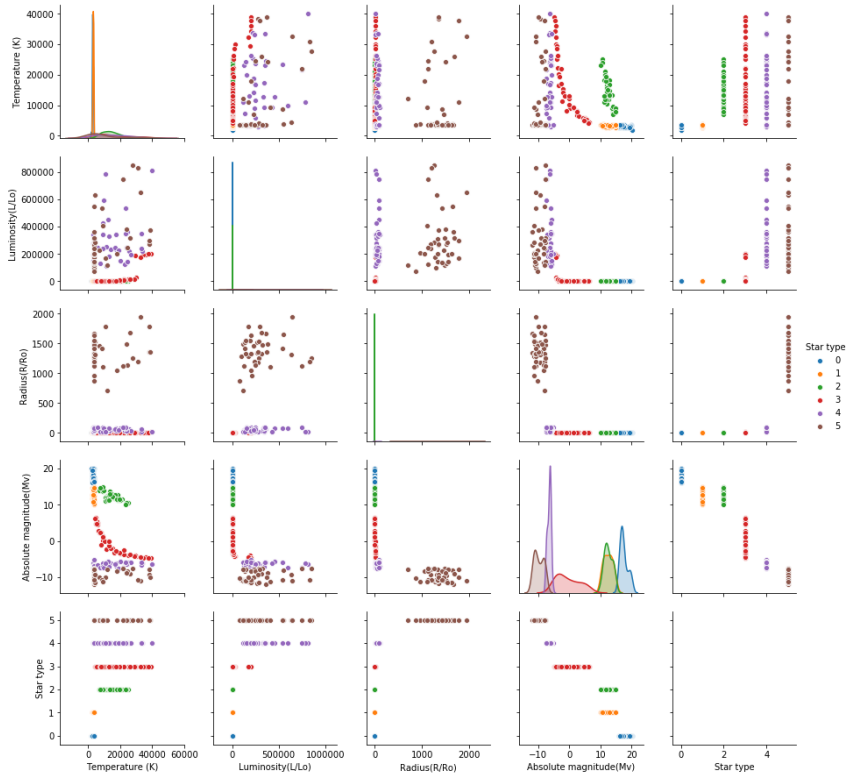
- Density plot



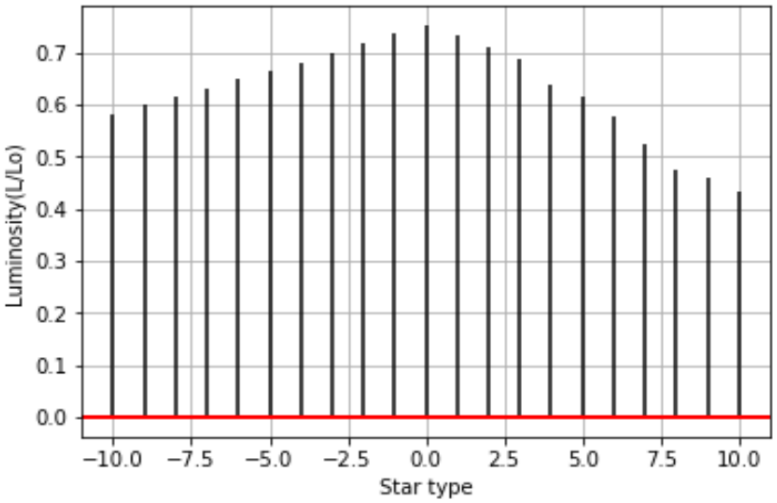
- Box plot



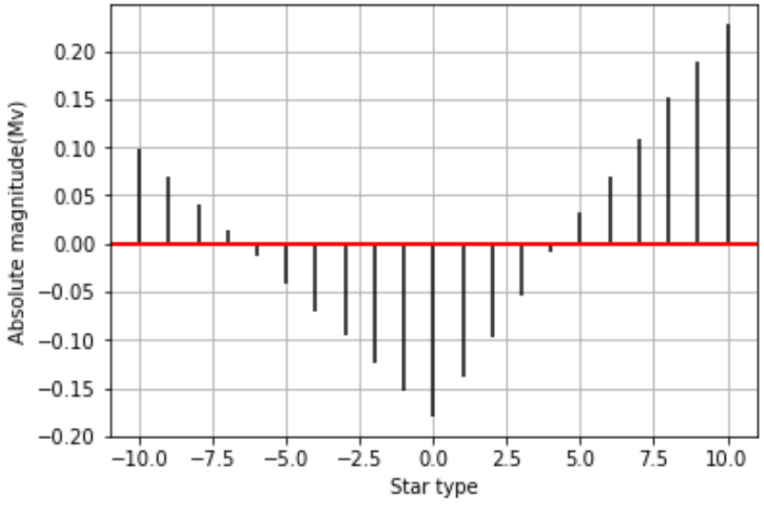
- Pair plot



- Corlogram for Luminosity



- Corlogram for Absolute magnitude



70

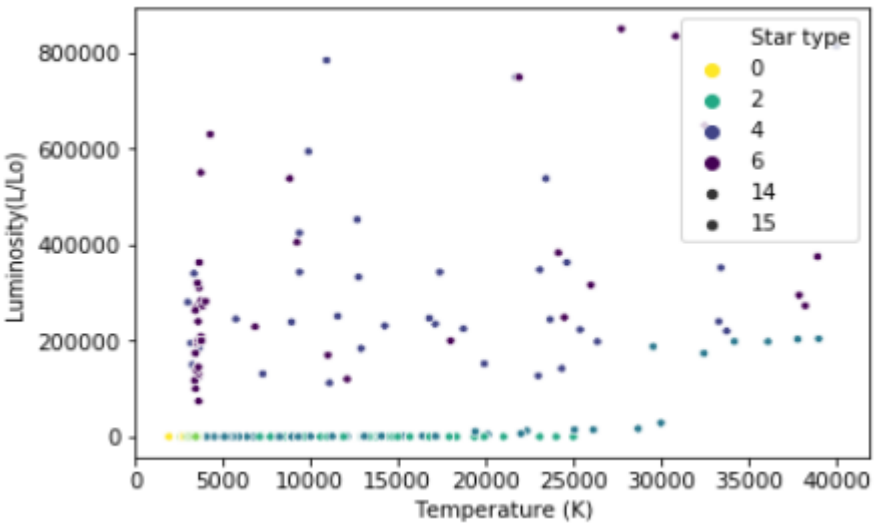
3.5. Machine Learning models

In this paper, different classifiers have been created which are:

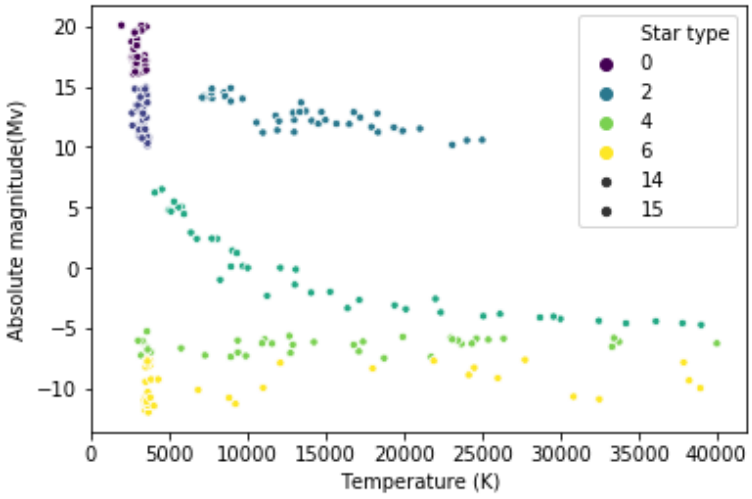
- Gaussian NB
- Bernoulli NB
- Stochastic Gradient Descent
- K-NN
- Logistic Regression
- Random Forest
- Decision Tree
- SVM

3.6. Hertzsprung-Russell diagram

- HR Diagram between Luminosity and Temperature



- HR Diagram between Absolute magnitude and Temperature



4. Result and Discussion

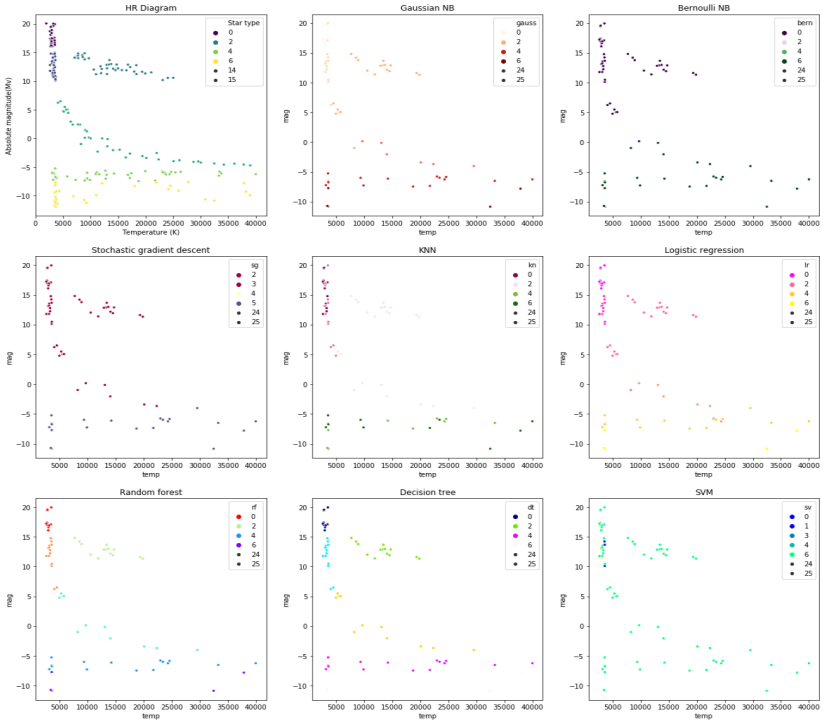
4.1. Performance evaluation table

Algorithm	**Accuracy**	**Precision**	**F1-score**	**Recall**
Gaussian NB	0.833	0.863	0.824	0.833
Bernoulli NB	0.236	0.056	0.091	0.236
Stochastic Gradient Descent	0.304	0.035	0.059	0.180
K-NN	0.597	0.656	0.601	0.597
Logistic Regression	0.833	0.844	0.821	0.833
Random Forest	0.972	0.975	0.972	0.972
Decision Tree	0.972	0.975	0.972	0.972
SVM	0.111	0.185	0.077	0.111

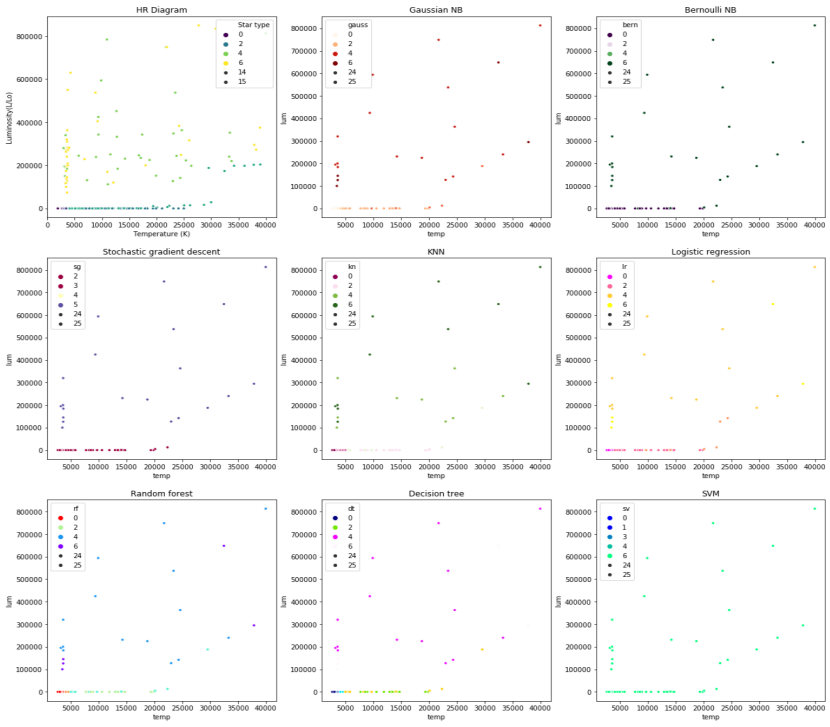
Table 1: Performance of Classifiers

4.2. HR Diagram w.r.t predicted Star type

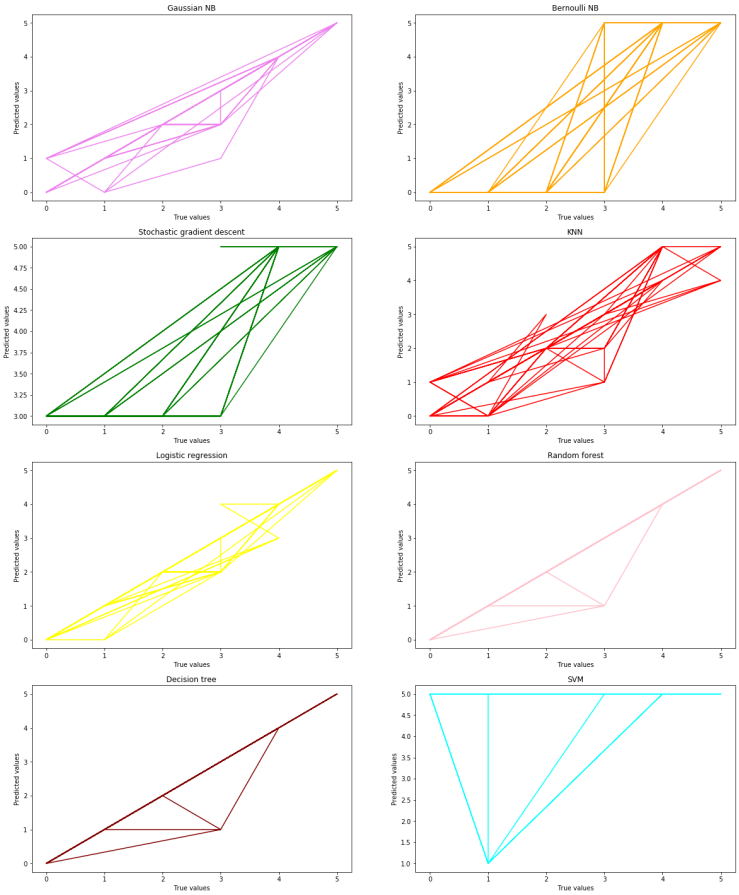
- HR Diagram between Absolute magnitude and Temperature w.r.t predicted Star type



- HR Diagram between Luminosity and Temperature w.r.t predicted Star type

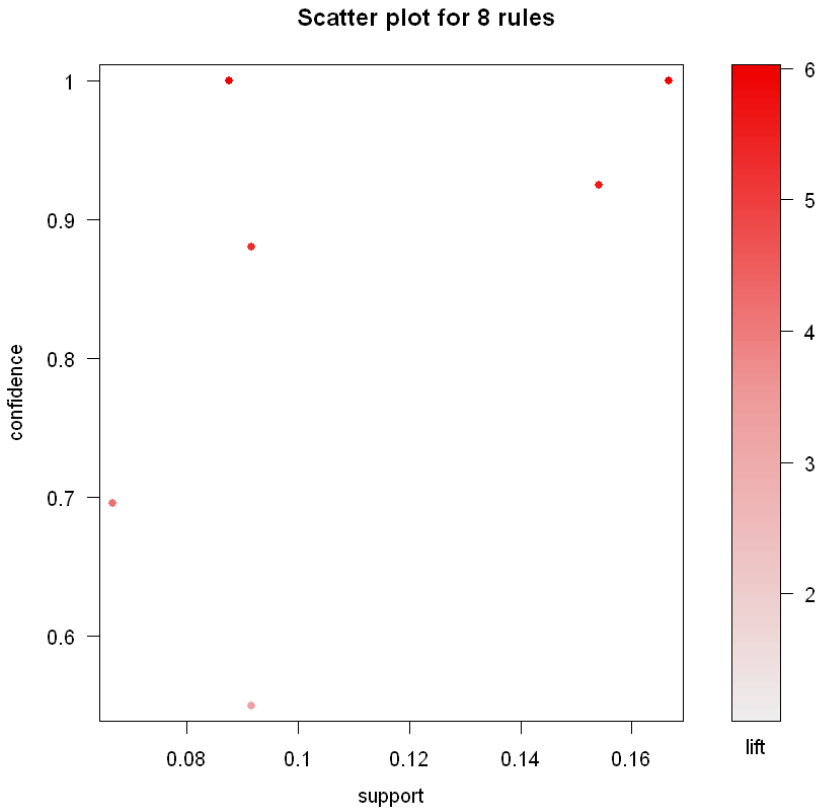


95 *4.3. Predictions*

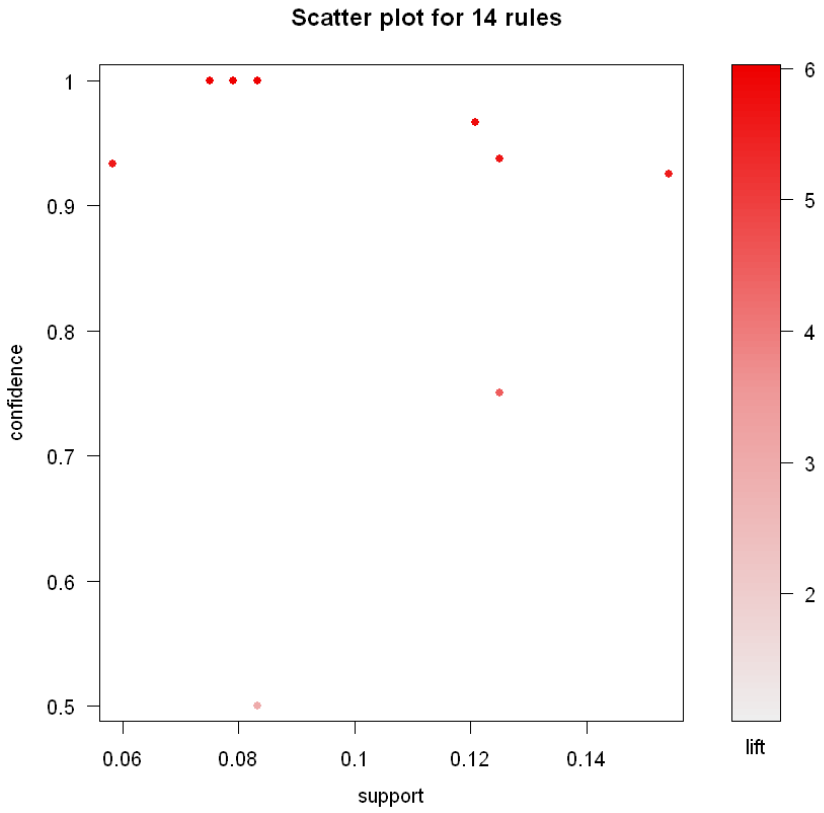


4.4. Association rules

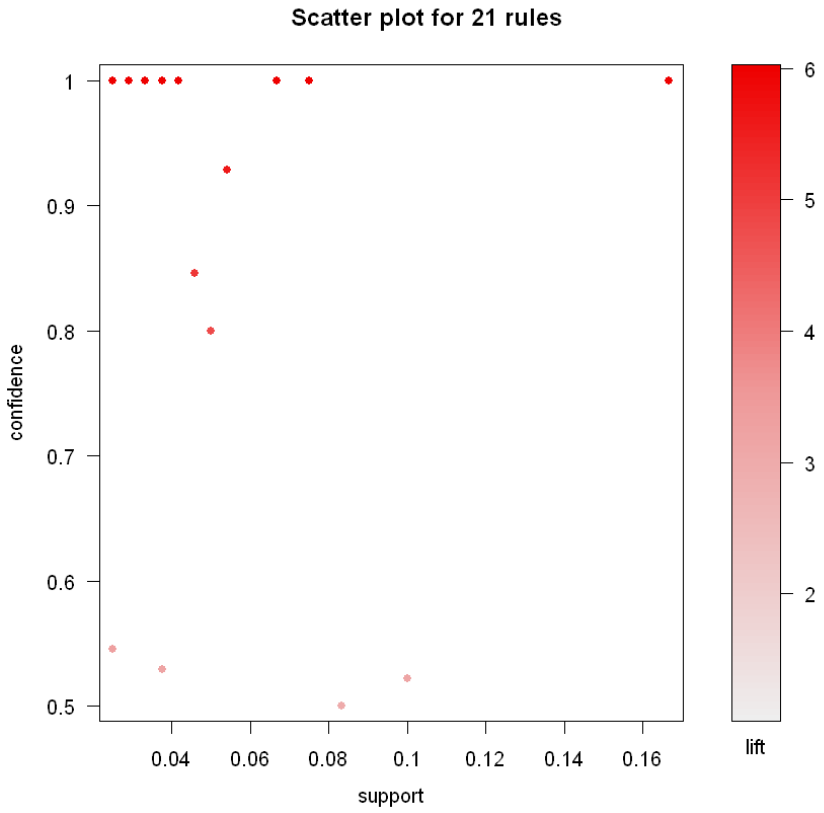
- Star type 0



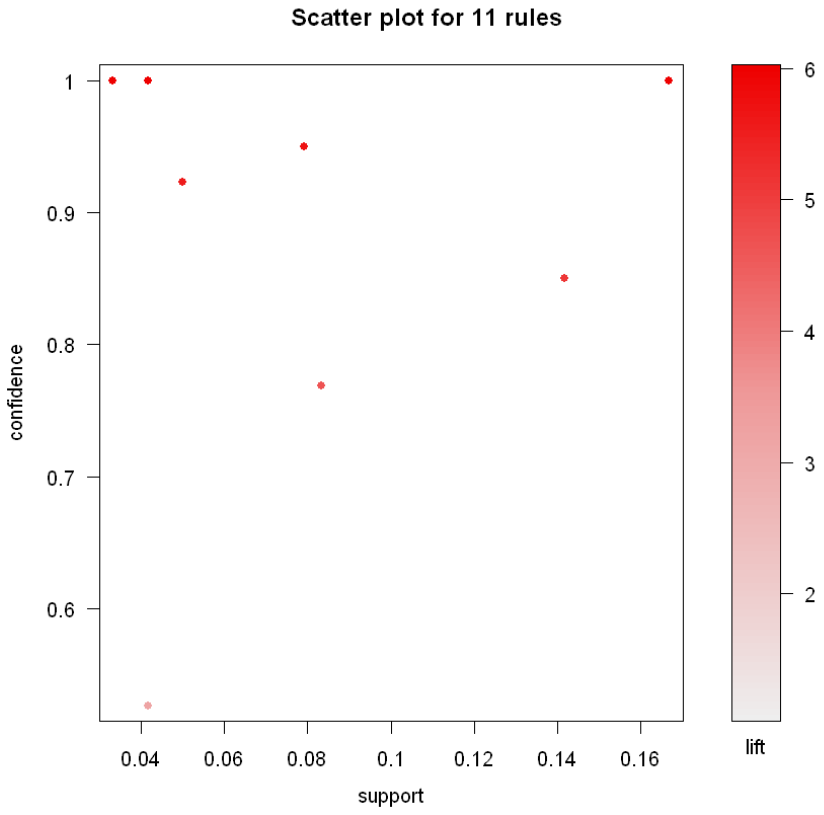
- Star type 1


Scatter plot for 14 rules
confidence
1
0.9
0.8
0.7
0.6
0.5
0.06
0.08
0.1
0.12
0.14
support
lift
6
5
4
3
2

- Star type 2

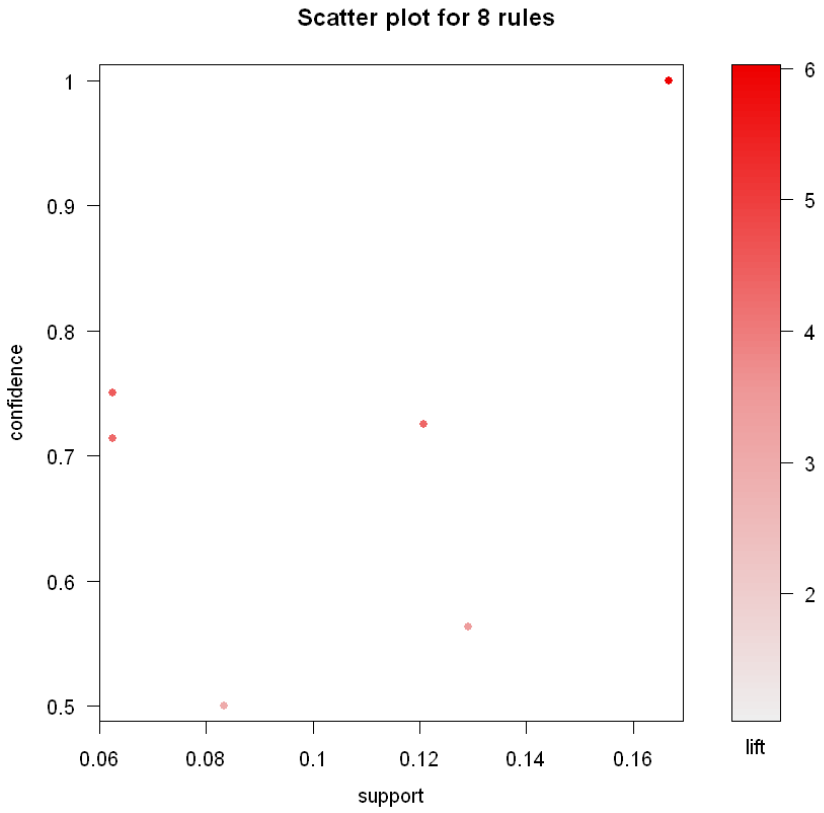

Scatter plot for 21 rules
confidence
1
0.9
0.8
0.7
0.6
0.5
0.04
0.06
0.08
0.1
0.12
0.14
0.16
support
6
5
4
3
2
lift

- Star type 3

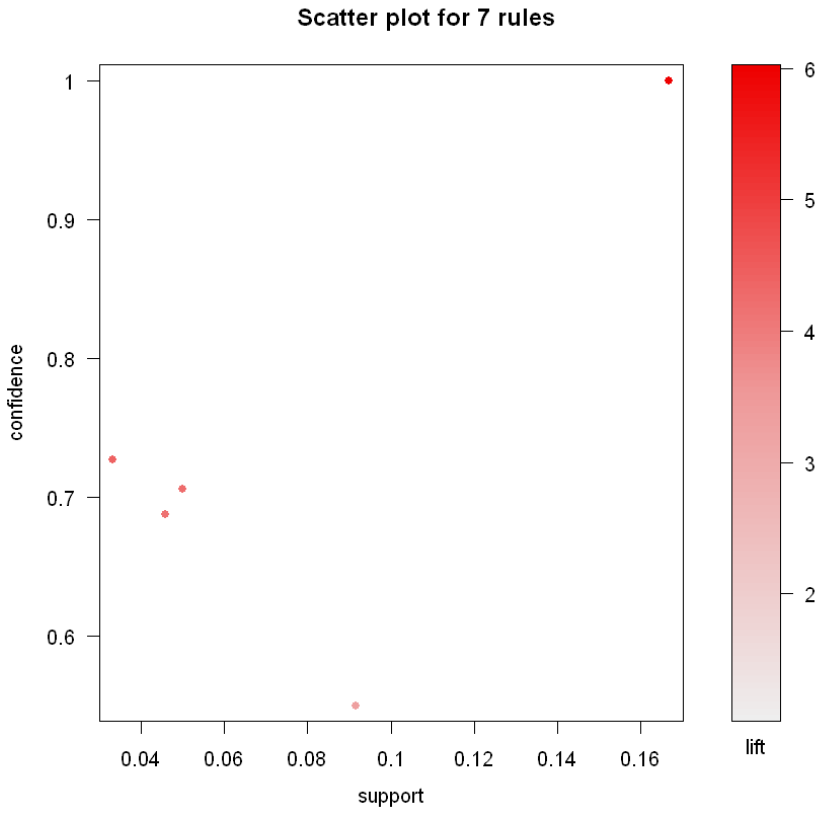


105

- Star type 4


Scatter plot for 8 rules
confidence
1
0.9
0.8
0.7
0.6
0.5
0.06
0.08
0.1
0.12
0.14
0.16
support
lift
6
5
4
3
2

- Star type 5


Scatter plot for 7 rules
confidence
1
0.9
0.8
0.7
0.6
0.04
0.06
0.08
0.1
0.12
0.14
0.16
support
lift
6
5
4
3
2

[1, 2, 3, 4, 5, 6**? ?** , 7]

https://www.nationalgeographic.com/science/space/universe/stars/

https://science.nasa.gov/astrophysics/focus-areas/how-do-stars-form-and-evolve

References

[1] C. Corbally, R. Gray, An expert system for classifying stars on the mk spectral classification system, The Astronomical Journal 147 (01 2013). doi:10.1088/0004-6256/147/4/80.

[2] G. Ledrew, The Real Starry Sky, 95 (2001) 32.

[3] C. Mackenzie, K. Pichara, P. Protopapas, CLUSTERING-BASED FEATURE LEARNING ON VARIABLE STARS, The Astrophysical Journal 820 (2) (2016) 138. doi:10.3847/0004-637x/820/2/138.
URL https://doi.org/10.3847%2F0004-637x%2F820%2F2%2F138

[4] S. Schmeja, D. A. Gouliermis, R. S. Klessen, THE CLUSTERING BEHAVIOR OF PRE-MAIN-SEQUENCE STARS IN NGC 346 IN THE SMALL MAGELLANIC CLOUD, The Astrophysical Journal 694 (1) (2009) 367–375. doi:10.1088/0004-637x/694/1/367.
URL https://doi.org/10.1088%2F0004-637x%2F694%2F1%2F367

[5] K. L. Adelberger, C. C. Steidel, M. Pettini, A. E. Shapley, N. A. Reddy, D. K. Erb, The spatial clustering of star-forming galaxies at redshifts 1.4 z 3.5, The Astrophysical Journal 619 (2) (2005) 697–713. doi:10.1086/426580.
URL https://doi.org/10.1086%2F426580

[6] V. Bromm, R. B. Larson, The first stars, Annual Review of Astronomy and Astrophysics 42 (1) (2004) 79–118. arXiv:https://doi.org/10.1146/annurev.astro.42.053102.134034, doi:10.1146/annurev.astro.42.053102.134034.
URL https://doi.org/10.1146/annurev.astro.42.053102.134034

[7] M. V. Zombeck, Handbook of Space Astronomy Astrophysics, 2nd edition, Cambridge University Press, 1982,1990.