ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 2, Issue 4, April 2017

Space Object Classification And Redshift Estimation Using Astronomical Spectroscopic Data And Its Representation

[1] Prakritidev Verma, [2]Hemangini Malhotra, [3]Amit Kushwaha, [4]Rajat Ratra
[1] [2] [3] [4]Department of Computer Science, Bharati Vidyapeeth's College of Engineering, New Delhi-110063, India

Abstract: ***The paper proposes a mathematical model for automated classification of space objects and redshift value estimation by applying the Sloan Digital Sky Server (SDSS) Dataset. Redshift is a measure of how fast a celestial object is moving relative to us, which is a basis for universe expansion. Furthermore, this paper is an attempt at Universe mapping in 2-Dimension as well as an interactive 3-Dimensional View.***

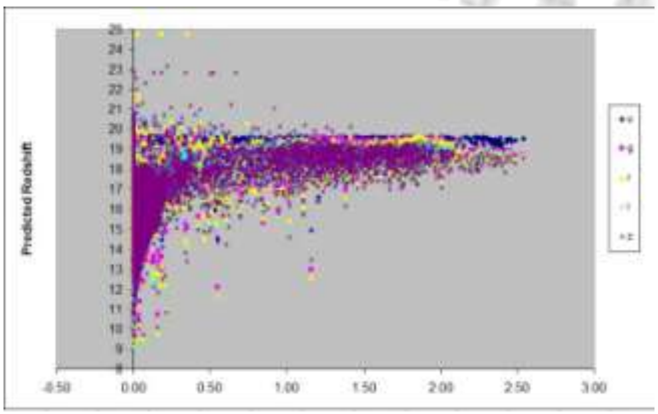
Keywords— Classification, Hubble's Law, K Nearest Neighbours, Random Forest, Redshift, Space Objects, 2D mapping, 3D mapping of universe.

I. INTRODUCTION

Hubble's law states a direct relation between the distance to a galaxy and its recessional velocity as determined by the redshift. It focuses on the idea of the expanding universe at a constant speed per unit distance. Thus, objects that are more distant move faster than nearby ones [15]. The slope of the relation, Ho, is the Hubble Constant; it represents the constant rate of cosmic expansion caused by the extension of space-time itself. Although the expansion rate is constant in all directions at any given time, this rate changes with time throughout the life of the universe [8].

$$\nu = H_o D \qquad -(1)$$

Where, ν represents Recessional velocity which is defined as the rate at which an astronomical object is moving away, H_o = Hubble Constant, D = Distance



The redshift is an astronomical term that describes the change in the atom spectra towards the red end of the spectrum when compared with a laboratory standard on earth [7]. Mathematically, the redshift, z, is defined as the calculated change in wavelength, when compared with the standard, divided by that laboratory standard wavelength.

$$z = \Delta\lambda/\lambda \qquad -(2)$$

If the change in wavelength is given by Δλ and the laboratory standard wavelength is represented by λ, then the redshift is defined as

$$\frac{v}{c} = \frac{1-(1+z)^2}{-1-(1+z)^2} \qquad - (3)$$

Where, v/c represents distance metric and z represents the numerical representation of redshift value. This value has a great significance in the mapping of galaxies and quasars. The more the value of z, far is the space object and more shifted it is towards the red end of the spectrum. Thus, it will appear redder to the observer as shown in Figure 1.

Fig. 1. Representation of Redshift with u-g-r-i-z spectra

The use of spectroscopic data for estimation is of significance here. The difference in features used for photometric analysis [13] of data may lead to varied results and thus, a spectrum-based data acts as a standard. Secondly, due to the enormity of celestial objects, it is highly computational to estimate the features through image-based data and shifts the focus of the research from redshift estimation and representation.

II. RELATED WORK

Zhang et al. [12] utilised various supervised and unsupervised methods for classification purpose for the space objects. The paper proposes Support Vector Machine to be highly accurate.
Hanula et al., [5] proposed an immersive visual mining and analysis tool for cosmological data, which allows domain experts to interact with visual representations of spatial and non-spatial cosmology data.
Benitez et al., [13] used Bayesian statistics for Redshift estimation from photometric data overcoming the "training set" issues.
The development of fast and accurate methods of photometric redshift estimation is a vital step towards being able to fully utilize the data of next-generation surveys within precision cosmology. Freeman et al., [4] applied a specific approach to spectral connectivity analysis (SCA) called diffusion map applying regression to make redshift predictions.

III. METHODOLOGY

3.1 DATA COLLECTION

Data Collection involves extracting the Dataset from SkyServer Search [11] using SQL Query Generation Method. The dataset consists of 5 million data points consisting of Celestial Coordinates: Right Ascension, like Celestial Longitude and Declination like Celestial Latitude for describing positional coordinates. Other Spectral data such as ultraviolet, green, red, infrared, zInfrared and numerical value of Redshift, i.e., Z-factor. Each row of data corresponds to an ObjectID describing different galaxies and quasars in the universe. It also contains a column for class for galaxies, quasars and stars with numeric representation of data.

3.2 DATA PREPROCESSING

In order to get the dataset ready, it is essential to understand that only spectrum based features such as ultraviolet, green, red, infrared and zInfrared values were used for Redshift estimation.
Initially, it is important to remove any outliers from the dataset as it may affect the accuracy of the results.
To make the classification step easier, the categorical features are transformed to numeric data for the class field of the dataset. This step is necessary for the proper representation of the distinct elements of the variable.

Another point to be noted being that Right Ascension (RA) acts and Declination (DEC) act as a positioning coordinates for the space elements which can be used for universe mapping.

3.2 CLASSIFICATION

This step requires classification of space objects into Galaxies, Quasars and Stars. Since, targeting specific space objects has always been an area of research in scientific industry; this paper proposes an optimised K-Nearest Neighbors (KNN) Classifier.

KNN is an instance based supervised learning technique. The KNN classifier alone gave poor results in the classification of space objects thus; the dataset was cross-validated to get a better result out of the SDSS dataset. This is known to be the Optimised KNN classifier.

3.4 REGRESSION

Regression is useful for Redshift estimation using the U, G, R, I, Z-spectra. The factor being evaluated under the Redshift estimation is the numerical 'z' factor which tells how far the space object lies.

A Random Forest Regressor creates subsamples of the dataset and uses averaging to improve the predictive accuracy.

In Figure 2, the workflow of the above-described process is given.

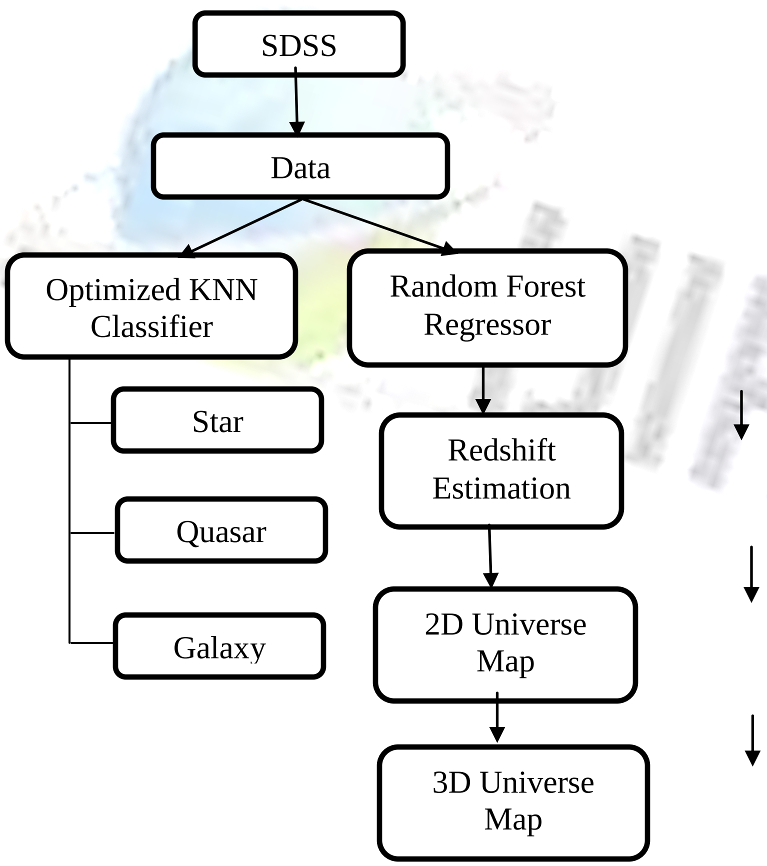


Fig. 2. Flowchart for Object Classification and Redshift Estimation

3.4 2-D MAPPING

The 2-Dimensional Mapping of the universe has been done using the Sloan Digital Sky Survey (SDSS) dataset [11] using 1 million points. The Celestial Coordinates: Right Ascension and Declination have been put to use for placing of the coordinates. Using the value of redshift's Z-factor, the density of points is shown with varying colour densities.

IJIRMET ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 2, Issue 4, April 2017

Right Ascension acts and Declination act as a positioning coordinates for the space elements and are here converted into degrees for better visualization.

3.5 3-D MAPPING

Two-Dimensional map visualizes the universe in an appropriate manner but has a major drawback of overlapping of space objects due to the absence of third dimension, i.e., they completely disregard the depth coordinate (Z-coordinate). Thus, to be able to accurately visualize the placement of galaxies and quasars, 3 Dimensional mapping is necessary.
Here, the third dimension being considered is the Redshift through the formula:

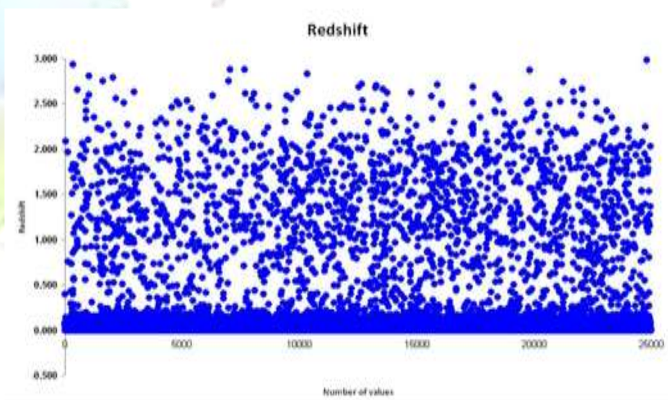
$$z = \frac{\lambda}{\lambda o} \qquad \text{- (3)}$$

Where, λ is the measured wavelength and λ_o is the rest wavelength.

IV. RESULTS

Upon using multiple classifier algorithms as shown in Table 1, it is noted that Support Vector Machine and Optimised K-Nearest Neighbours (KNN) perform better than the rest. Choosing Optimised KNN as the suitable algorithm is a result of the slow processing speed of SVM, i.e. Support Vector Machine.

The KNN classifier with cross-validated dataset produces an accuracy of 96.02%.



Classifier		Precision	Recall	Accuracy
Logistic Regression	Star	83%	71%	83.54%
	Quasar	90%	78%	
	Galaxy	83%	92%	
Decision Tree Classifier	Star	76%	57%	77.91%
	Quasar	86%	72%	
	Galaxy	77%	92%	
Support Vector Machine	Star	93%	95%	94.13%
	Quasar	91%	85%	
	Galaxy	96%	95%	
Optimised KNN	Star	95%	96%	96.02%
	Quasar	95%	92%	
	Galaxy	97%	96%	

Table 1. Classification results

Similarly, upon carefully examining as shown in Table 2, the results from different algorithms for redshift estimation using

U, G, R, I, Z-spectra, we have seen that Random Forest Regressor outperforms from the rest giving an accuracy of 76.40%.

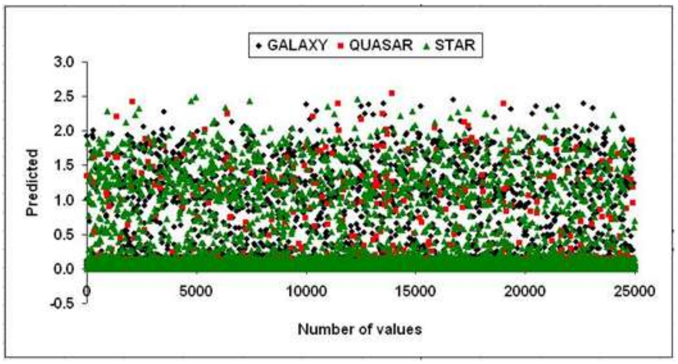


Fig. 3. Visualization of classification of space objects

Fig. 3 represents the result of classification of space objects into Galaxy, Star and Quasar graphically.
The precision factor is defined for the correctly predicted value by the model whereas; recall is defined as how sensitive the model is in classifying the object presented to the model.

Algorithm	Accuracy
Adaboost Regressor	60.08%
Gradient Boosting Regressor	51.07%
Decision Tree Regressor	57.90%
Random Forest Regressor	76.40%

Table 2. Redshift Estimation results

The performance of Random Forest Regressors is justified by the fact that the dataset consists of weak learners and specific association rules cannot be applied. This why Decision tree regressor doesn't show significant results. Random forest takes a multitude of decision trees and evaluates the dataset and Redshift estimation carefully. Thus, the shortcomings of the dataset are somewhat overcome by this algorithm.

Fig. 4(a). Observed redshift value graphical representation

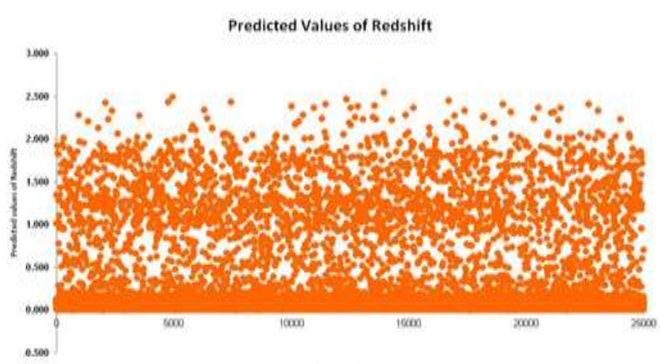


Fig. 4(b). Predicted redshift value graphical representation

Figure 4(a) represents the observed value of redshift from the SDSS Dataset. Figure 4(b) represents the result of the regression from u-g-r-i-z spectra to estimate the value of redshift.

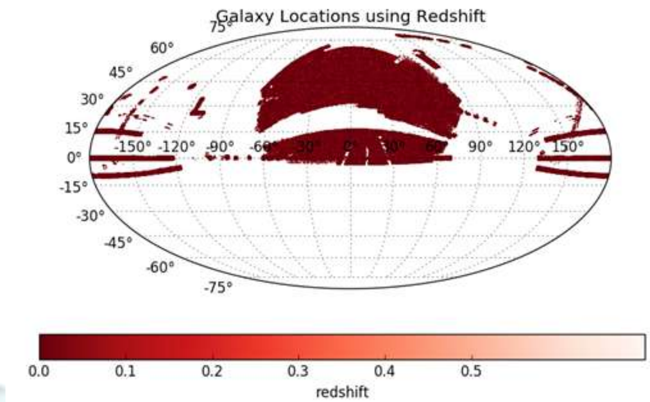


Fig. 5. 2-Dimensional Representation of Universe

Figure 5 represents the 2-Dimensional Mapping of the universe based on Redshift. The values of redshift vary from 0 to 7. As seen from the figure, more the value of redshift, more the elements lie away from the observing point, i.e., Earth.

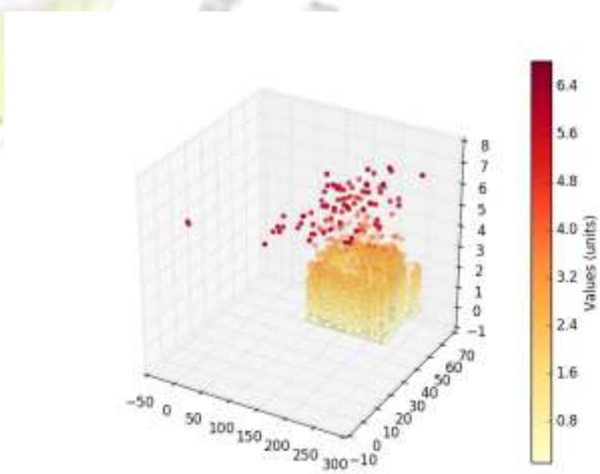


Fig. 6(a). 3D mapping of the Universe based on redshift

IJIRMET ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 2, Issue 4, April 2017

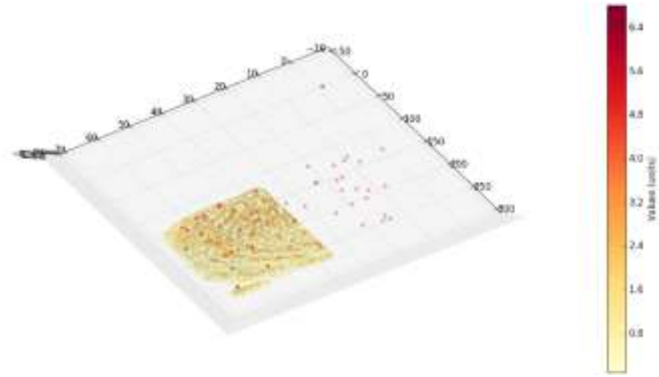


Fig. 6(b). 2D to 3D representation

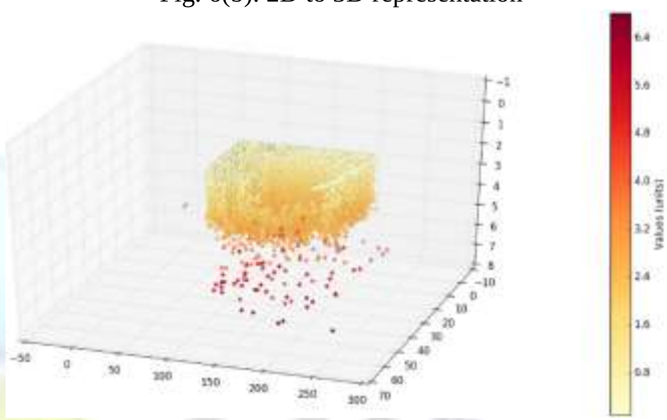


Figure 6(c). Inverted map representation based on redshift

Figure 6 represents 3D mapping of the universe. The color bar represents the scale of redshift. It is evident from the figure that the redder the data points are, the farther the data point is from the reference point. The data points represent Galaxies, Stars and Quasars.

5 CONCLUSION

Various approaches have been applied in the task of separating the space objects into Galaxies, Quasars and Stars. A cross-correlated KNN classifier shows the best results with a considerably high processing speed with 5 million data points. Similarly, the regression gives significant results with a Random forest algorithm technique.

Due to the limitation of unavailability of an adequate public dataset, the training of Mathematical Model doesn't yield highly accurate results. However, this model can be used for a better understanding of the universe through mapping by using the redshift and celestial coordinates as shown in the paper.

The 2-Dimensional mapping of the universe is not sufficient to visualize the universe due to overlapping of space objects, which is why an interactive 3-Dimensional view has been created for better redshift representation.

ACKNOWLEDGEMENT

We thank Mr. Varun Srivastava, Assistant Professor, Computer Science Department, Bharati Vidyapeeth's College of Engineering, Delhi for assistance with carrying out the research project. His comments and suggestions greatly improved the manuscript. His great insights into the Machine Learning problem dealt with and his belief in us for completion of the project helped us throughout the course of the research.

REFERENCES

[1] Steenstrup Pedersen, Kim, et al. "*Shape index descriptors applied to texture-based galaxy analysis*." Proceedings of the IEEE International Conference on Computer Vision. 2013.

[2] Pizarro, Daniel, et al. "*Clustering of 3D spatial points using maximum likelihood estimator over voronoi tessellations: Study of the galaxy distribution in redshift space*." Voronoi Diagrams in Science and Engineering, 2006. ISVD'06. Third International Symposium on. IEEE, 2006.

[3] Luciani, Timothy Basil, et al. "*Large-Scale Overlays and Trends: Visually Mining, Panning and Zooming the Observable Universe*" IEEE transactions on visualization and computer graphics 20.7 (2014): 1048-1061.

[4] Freeman, P. E., et al. "*Photometric redshift estimation using spectral connectivity analysis*" Monthly Notices of the Royal Astronomical Society 398.4 (2009): 2012-2021.

[5] P. Hanula, K. Piekutowski, C. Uribe, A. Nishimoto, K. Almryde, J. Aguilera, and G. E. Marai. "*Cavern halos: Exploring spatial and nonspatial cosmological data in an immersive virtual environment*" In Virtual and Augmented Reality, 3D, and Stereoscopic Systems Conference Poster Compendium, Electronic Imaging'16, 2016.

[6] R. S. J. Kim, J. V. Kepner, M. Postman, M. A. Strauss, N. A. Bahcall, J. E. Gunn, R. H. Lupton, J. Annis, R. C. Nichol, F. J. Castander, J. Brinkmann, R. J. Brunner, A. Connolly, I. Csabai, R. B. Hindsley, Z. Ivezi´c, M. S. Vogeley, and D. G. York. "*Detecting Clusters of Galaxies in the Sloan Digital Sky Survey"* I. Monte Carlo Comparison of Cluster Detection Algorithms. The Astronomical Journal, 123:20–36, January 2002.

[7] Madau, Piero, et al. "*High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to z∼ 4*." Monthly Notices of the Royal Astronomical Society 283.4 (1996): 1388-1404.

[8] Linder, Eric V. "*Exploring the expansion history of the universe*" Physical Review Letters 90.9 (2003): 091301.

[9] Bromm, Volker, and Abraham Loeb. "*The expected redshift distribution of gamma-ray bursts*" The Astrophysical Journal 575.1 (2002): 111.

[10] Wu, Xue-Bing, et al. "*An ultraluminous quasar with a twelve-billion-solar-mass black hole at redshift 6.30*" Nature 518.7540 (2015): 512-515.

[11] "*SkyServer DR7 Search Form*" SkyServer DR7 Search Form. N.p., n.d. Web. 11 Mar. 2017.

[12] Zhang, Yanxia, et al. "*Classification of Quasars and Stars by Supervised and Unsupervised Methods*" International Astronomical Union. Proceedings of the International Astronomical Union 8.S288 (2012): 333.

[13] Benitez, Narciso. "*Bayesian photometric redshift estimation*" The Astrophysical Journal 536.2 (2000): 571.

[14] Kirshner, Robert P. "*Hubble's diagram and cosmic expansion*" Proceedings of the National Academy of Sciences 101.1 (2004): 8-13.

[15] Milani, A., et al. "*Correlation and orbit determination of space objects based on sparse optical data*" Monthly Notices of the Royal Astronomical Society 417.3 (2011): 2094-2103.

BIOGRAPHIES

Prakritidev Verma is currently pursuing B.Tech in Computer Science Engineering at Bharati Vidyapeeth's College of Engineering. His main areas of interest are Deep Learning, Reinforcement Learning, and Natural Language Processing. His current publications include, "Bitcoin Value Analysis Using Graphical Visualization Technique", IEEE Xplore Digital Library.

Hemangini Malhotra is currently pursuing B.Tech in Computer Science Engineering at Bharati Vidyapeeth's College of Engineering. Her main areas of interest are Image Processing and Machine Learning. Her current publications include A Novel Method for Fingerprint Recognition Using Trifurcation, Eye Enclosure" at International Journal of Control Theory and Applications, 2016. She is currently a member at Computer Society of India (CSI).

Amit Kushwaha is currently pursuing B.Tech in Computer Science Engineering at Bharati Vidyapeeth's College of Engineering. His main areas of interest are Robotics, Computer Vision Data Science. He is currently a member at IEEE and IEEE Computer Society.

Rajat Ratra is currently pursuing B.Tech in Computer Science Engineering at Bharati Vidyapeeth's College of Engineering. His main areas of interest are Data Mining and Image processing. He is currently a member at Computer Society of India (CSI).