Literature Review of Deep Learning Applied to Astronomy Data Analysis

Engler, S.T., University of Hawaii, Department of Information and Computer Science
12th November 2019

1. Introduction

Astronomy is going through a paradigm change where the data collected by ground and space telescopes represents huge amounts of data, significantly more than seen before. Deep Learning (DL) has started to have an impact on Astronomy and Astrophysics through analysis of these large data sets. This literature review looks at the foundations of Deep Learning literature and then take a look at how and where these DL methods are applied in astronomy and astrophysical data analysis. Deep Learning breaks into two main types of data processing in astronomy. First, image processing through Convolutional Neural Networks (CNN), which allow for a great number of astronomy applications such as galaxy age categorization, and stellar spectral classification. Second through time series analysis, which allows for DL to be applied to eclipsing exoplanet identification through space telescope data such as Kepler, TESS, and ground-based observations using measurements of radial velocity extracted from spectral analysis. Deep Learning is in its infancy making its way into astronomy data. This literature review found that Deep Learning is being applied in a variety of different discipline in astronomy and finds that there are many research areas that seem to be untouched thus far. The literature also reveals astronomers are looking towards deep learning to be developed as the main tool for handling the large amount of data from the next generation of telescopes. The review concludes by identifying potentially untapped areas of research which could be pursued for future research activities in Astronomy and Astrophysics. [1]

2. Deep Learning

Deep Learning is a subset of Machine Learning, which is also a subset of Artificial Intelligence. Deep Learning uses techniques to learn multiple levels of representation within a very large data set. Dominantly when people speak of deep learning they are speaking about deep artificial neural networks, rather than machine learning or artificial intelligence. [2] Deep learning models need large scale data sources in which traditional NN would overfit after a certain amount of time. Overfitting is when the NN will output the same solution as a result of over training the system. [3] Deep Learning has outstanding success in almost every scientific venture it is applied to. [2] DL breaks into three different general approaches of supervised, reinforcement, and unsupervised learning. [2] Supervised learning uses labelled data that compares a loss value to training data and iteratively modifies the neural network parameters for better results. Reinforcement Learning is based on partially labelled data sets. Unsupervised learning has no data labels and learns the important features of the dataset and will discover structure within the data. [2] Google created 'Deep Mind' which is the first unsupervised learning system. [4, 5] The challenge with RL is that there is no solution to calculate a loss function from. If there are few enough parameters to tune then parameter optimization can be done though methods such as cross entropy calculations, and simulated annealing. [2] Types of Deep supervised learning are Deep Neural Network (DNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) which includes the Long-Term Short Memory (LSTM) and Gated Recurrent Units (GRU). [2] Deep Learning is built off Neural Networks, and the Deep aspect did not appear until 2006. Initially, back in 1943 it was shown that a perceptron could be used to solve such problems. Later, it was shown that a Turning machine could be created using neurons. [6] The building blocks of neural nets are the perceptron. The perceptron is a single neuron. Let us take m pixels as input. Multiplying each input by weights of theta and taking their sum, the result is passed through a activation function that produces output y. [3] Perceptron were able to come to a solution so long as one was able to guide towards a solution state. However, the perceptron is unable to solve XOR logical problems and the research was stopped. It was not until 1985 that the backpropagation algorithm was introduced which created more usable NN. [7] After the introduction of back propagation, convolutional neural networks appeared [8] in 1998. Following this, Deep Neural Networks appeared with [9,10] From there, the first full DNN appeared in 2012 [11]. The primary components of the NN are the neurons which are governed by activation functions, normally a sigmoid function. NN is connected in multiple layers with at least one hidden layer. As the NN computes, the NN has weights between the connections between the neurons the weights are found though iterating through the NN and comparing to the objective function (solution). The weights are adjusted using gradient decent which is an optimization algorithm. Once a local minimum is found, the NN will use back propagation to push the new information back through the NN system. For DNN, often a stochastic gradient decent is used to prevent the system from getting stuck in a local minimum. [12] The learning rate is the step size of the training which can speed up the training process. It can cause the system to diverge if the steps are too large, or converge quickly if the steps are the right size. As the NN starts to converge, often the learning rate will be reduced at a constant or exponential decay. [2]

3. The Convolutional Neural Network

Fukushima was the first person to propose the CNN, which was able to recognize hand writing of numbers. [13] The CNN in its simplest form consists of a feature extractor and a classification. The CNN starts with a convolution layer, a max-pooling layer, and a classification layer. [3] The CNN has a max pooling layer that allows the network to process shape variations. This layer prevents the NN form having too many parameters to control. [3] The max pool layer allows the CNN to be less sensitive to the gradient problem and ultimately produces a highly accurate and efficient NN. [3] The classification layer is often a feed-forward NN. [9,14] CNN take advantage of the image inputs allow the architecture to constrain in an efficient way. When input is passed to the convolution layer the output is an activation map that extracts features from the input data to pass further. Pooling layers further reduce the number of parameters. [3] The convolution layer consists of feature maps that go through kernels which are passed into activation functions such as the sigmoid, SoftMax, or rectified linear) which then create feature maps. These feature maps can then be combined with multiple input feature maps. [2] The max-pooling layer is really a form of subsampling. This layer down samples the output from the convolution layer. Each output layer is down sampled into half the size of its incoming dimension. The down sample is found through averaging the neighboring values, or by taking the max value. [2] The classification layer scores the features extracted from the convolution layer. This information is backpropagated through the CNN with the convolutional operation on the filters and their previous layer. [2] The output of the CNN is a fully connected layer that output is compared to the output layer for error generation and a mean square loss is found and then backpropagated through the system. [3] CNN are the first architecture to successfully solve into a deep network with multiple hierarchal layers. The number of parameters of the network are greatly reduced through the CNN topology through the sharing of parameters. This parameter sharing has the consequence of creating equivariance in the system. [15]

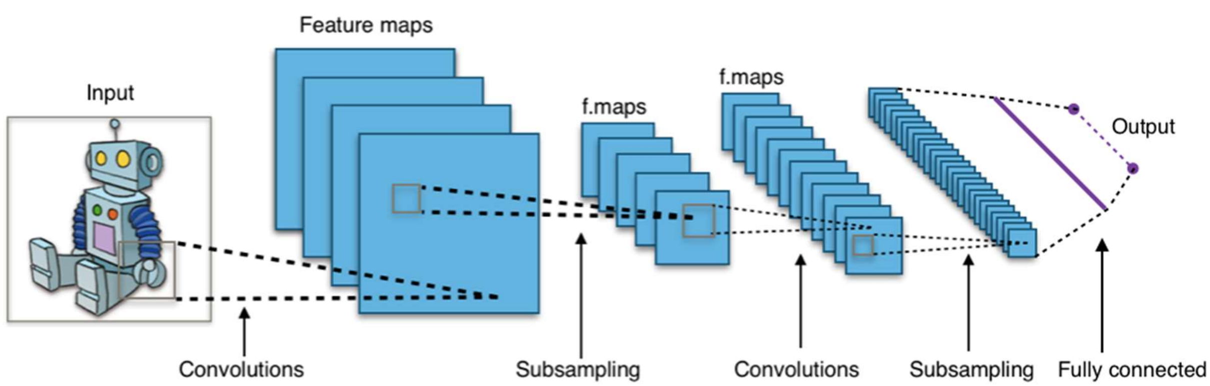

Figure 1: The CNN. Source: (https://en.wikipedia.org/wiki/Convolutional_neural_network)

4. Deep Recurrent Neural Network

The Recurrent Neural Network (RNN) is a NN where the output from previous step are fed as input to the current step. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence. An RNN remembers each and every information through time and it is easily applied to time-series data. Another form of the RNN is the Long-Short Term Memory RNN (LSTM-RNN) which is the network passing its results to another copy of the network.

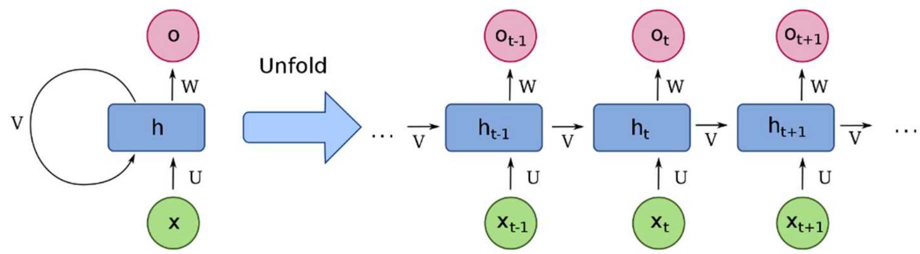

Figure 2: RNN Source: (https://en.wikipedia.org/wiki/Recurrent_neural_network)

The network can be a long string of time series data with the output being of equal length. [16] RNN by themselves are limited to the dependencies between each cell which is called the long-term dependency problem. To overcome this a LSTM cell is introduced to retain information indefinitely. [16] This is accomplished by using three different gates which will let the cell read or write information according to parameters that open or close these gates. They are cells with logistic and linear units where information gets into the cell when the "write" gate is open. It will stay in the memory until the "keep" gate is opened. The "read" gate will give access. It is also possible to back-propagate through the memory cells. [17] Each gate is simply a layer of neural nets governed by a sigmoid function. [18]

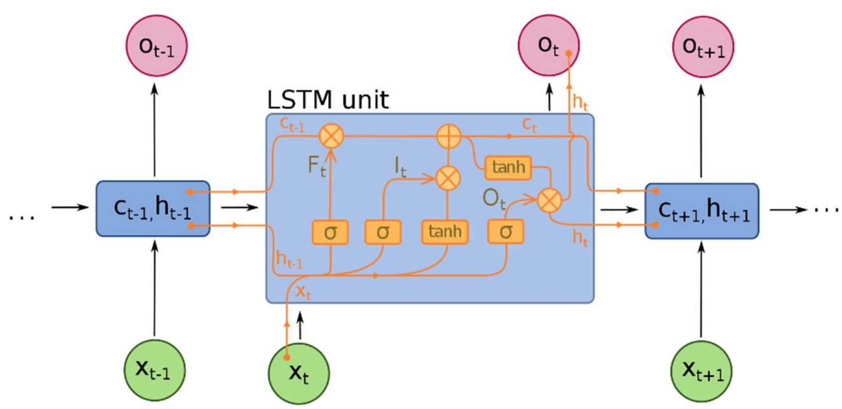

Figure 3: LSTM Cell (https://en.wikipedia.org/wiki/Recurrent_neural_network)

5. Deep Learning in Astronomy and Astrophysics Research

New generation telescopes such as the Atacama Large Millimeter Array (ALMA), Very Large Telescope (VLA) the sky survey telescope will produce so much data that processing this data automatically in the astronomy community is going to become a requirement. It is not enough to simply process the data, but the data needs to be capable of testing scientific questions based on the existing data. One research area in astronomy is the classification of astronomical sources. Classification is something that CNN are quite good at, however there are challenges in dealing with bias, mis-classified data, and portability of CNN to other datasets that it is not trained on.

5.1 Supervised Machine Learning

Supervised learning in astronomy can be seen in [29] where they employ a random forest for mapping the redshift of galaxies through spectroscopic analysis. They used 80K randomly selected samples from the data which yielded 600 trees for redshift estimation. Another group in [30] used an ANN to estimate photometric redshifts. They used the ANN to correlate the relation between the photometric measurements and the redshift for galaxies using the Sloan Digital Sky Survey (SDSS) data releases. The Sloan Digital Sky Survey consists of three-dimensional maps of the sky, and spectra for more than three million astronomical objects. [31] They used 30K galaxies where they broke the dataset up into 15K for training, 5K for validation, and 10K for evaluations. They used a committee of five networks being trained on these sets and then compared results for the evaluation.

Another group used machine learning methods to calculate start formation rates in galaxies in [32] a number of different models while also using the SDSS data. First, they used random forests Random Forest followed by an ANN trained under a quasi-Newton algorithm. They then they used k-fold cross validation and then selected features. The resulting Stellar Formation Rates from the ML produced are shown to be about within 6% error. It is felt that by improving the methods more they will be able to simulate an accurate catalog of SFR across the sky survey.

Supervised learning has also been used for applications in Cosmology where [33] discusses how Super Nova of type IA are a central role as cosmological candles making them prime targets in large-scale sky surveys. The challenge is that the Super Nova require classification through spectral imaging which is less available than photometric measurements. The challenge is to use ML to accurately classify the super nova using only photometric data. The challenge of this is the initial lack of data available for training. So, they used physical models to produce the photometric light curves for a variety of stars and super nova events with mixed results. With the incorporation of large sky surveys such as SDSS they expect to have a lot more data to incorporate for training and expect to produce more accurate results.

5.2 Unsupervised Machine Learning

Unsupervised classification relies on the statistical properties of objects while supervised classification is inherently bias through inconsistent human assessment. In the literature [19] classify variable stars and exoplanet transits using a Recurrent-CNN which uses the raw data to extract features. This allows for data analysis directly from raw data without tedious data preparation normally conducted through the supervision of humans. This research also simulates synthetic image sequences based on the instrumentation and observations to adapt the training of the RCNN. Using this simulation for training allows the system to be fine-tuned to the data from the actual experiment using only a small number of real data points.

Another application of unsupervised learning in astronomy was conducted by [34] where they used ML to identify diffuse interstellar bands. They used the algorithms to dive the dark interstellar bands into 250 strong spectra and six groups of weak diffuse interstellar dust. Using 1.5 million extragalactic spectra they were able to identify a new pair of weak absorption lines that have a strong correlation with dust extinction.

In galaxy type classification for large sky surveys research in [35] showed an unsupervised learning technique that classifies galaxies using only pixel data from images. They used no preselection or prefiltering of the data like in other ML application and created a method that can be used straight from the raw images. They used two ML algorithms and an image processing algorithm to group over 60K catalog galaxies using the shape, orientation, size of the galaxies and created a hierarchal cluster. Followed by using connected component labeling which creates subclassifications of objects. The resulting output assigns a label to each component in the cluster and sub clusters.

Research conducted in [35] used ML to try to deal with noisy photometric data. Using principal component analysis, they were able to classify noisy light curves and associate them with areas of the spectrum, allowing for regression of the redshift. This application allows for the detection of high red shifting quasars.

Another group used unsupervised learning as a form of outlier detection to discover abnormal galaxies, as seen in [36] through a random forest algorithm trained on both synthetic and real data. The data input is a matrix with each row being the spectrum of the galaxy and the columns being the features of the data, those being the light flux at each wavelength. The high resolution of the spectrum gives 15.7K features and over 2.3M objects. These are fed into a random forest decision tree which categorizes the features. The RF is trained with a mixture of both real a synthetic data. The results from this method are very good where the group finding galaxies with unusual kinematics, ones with abnormally high number of supernovae, and high ionization lines.

Unsupervised classification is seen in a number of papers based on structural features of astronomical objects such as galaxies. In [20] a CNN is used to classify galaxies into structural categories of spheroid, irregular, point source, or unclassifiable using a Hubble super novae survey. Complexities in the research arose with the accuracy of category labels used in the training process of the CNN due to the fact that the labels were picked out by humans with a bias towards what that person thought the category the galaxy belonged to. To reduce this bias the authors resorted to using five experts to classify a hundred randomly selected galaxies to help fine tune the CNN training.

A similar study by [21] classified galaxies through an online group of citizen-scientists as a form of creating a large categorized dataset. They then used this data set to train and compare to real data using a Quantile Regression Forest that provided a matrix of similarities between the image and training data.

Another study in unsupervised classification of galaxies was conducted by [22] using a large-scale dataset of over 30K images of galaxies for Deep K-means clustering and combining the data with feature selection. Amazingly the authors used only observed data from photometer, spectroscopy, and used no non-observational data. This approach was successful in identifying clusters of galaxies using different star formation history. This study showed that Deep Learning was successful in categorizing data with strictly physical data and alleviated label bias in categorical applications.

Deep CNN has also been applied to astrophysical simulations of stellar asteroseismology data. [23] It allowed the group to interpolate between different stages of stellar evolution within the complex evolutionary models. The space telescopes Kepler and TESS which are both photometric measurements of stars over a large section of the sky. are dominantly time-series datasets extracted from telescopes hunting for exoplanets. There is an extremely large amount of data that requires detection of a transit event within these data sets.

One of the key points of interest in the hunt for exo-planets is the hunt for Earth-like planets the detection of such transits is presence of red (correlated) noise in the light curves. Researchers in [24] Simulated the red noise through the application of hyperparameters and found that the noise can be filtered through by the RNN to extract data for earth like planets. Research conducted by [25] expanded upon an exoplanet DNN created by which had a 95% accuracy rate. Their work illustrates the importance of including expert domain knowledge in even state-of-the-art deep learning models when applying them to scientific research problems that seek to identify weak signals in noisy data. This research paved the way for Deep learning to be used as a major tool in astronomy and the hunt for exoplanets.

In Cosmology deep learning has been applied in research conducted by [26] developed a pixel-based approach to implement convolutional and pooling layers on the spherical surface estimation of cosmological parameter from simulated maps of the cosmic microwave background (CMB). show the applicability of our CNN to CMB parameter estimation using simple NN architecture, consisting of four convolutional and pooling layers. They were able to sample value temperature and polarization parameters directly from simulated maps using this method to better accuracy that simulations using Bayesian methods.

6. Discussion

This literature review examined the origins of Deep Learning at the basic level defining what convolutional neural networks are and the recurrent neural network in the LSTM form. The literature review then examined where DL has been influencing world of astronomy. It found that DL has only just begun to scratch the surface of astronomy, and has been used only lightly in a wide variety of astronomical applications. It also found that astronomers are looking toward DL to be the main tool to process data from the next generation of telescopes. The literature for DL in astronomy applications highlight Transfer Learning as one of the major concerns for future research. [27] Transfer learning is the ability to take data from one type of data and apply it to another type. For example, taking photometric data from TESS and applying it to spectroscopic data. Authors in the study [28] show were able to use a new approach to time domain adaptation of the problem. The literature also indicates additional problems within applying ML to astronomy. Namely learning algorithms are not finetuned to work with astronomical datasets as there is variation in the quality of astronomical data, and there currently no easy way to assess this within the DL algorithms. Specifically, astronomical measurements vary in terms of signal-to-noise ratios, measurement uncertainties, and simulations used in training have modelling uncertainties. The construction of classifiers that are able to handle situations such as taking a ML model trained on one dataset and applying it to another is a high priority. Additionally, the new telescopes will need to be able to train while the data is being collected. The literature [27] also calls for Deep CNN that are able present data in a manner that is scientifically interpretable and physically meaningful within the astrophysical models being tested.

6. References

[1] P. Hála, “Spectral classification using convolutional neural networks,” 2014.

[2] Z. Alom et al., “The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches,” CoRR, vol. abs/1803.0, 2018.

[3] I. Engineering and J. B. Raja, “SURVEY ON DEEP LEARNING ALGORITHMS,” vol. 5, no. 1, pp. 38–43, 2019.

[4] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529-533.

[5] Mnih, Volodymyr, et al. "Playing Atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602 (2013)

[6] McCulloch, Warren S., and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." The bulletin of mathematical biophysics 5.4 (1943): 115-133.

[7] Ackley, David H., Geoffrey E. Hinton, and Terrence J. Sejnowski. "A learning algorithm for Boltzmann machines." Cognitive science 9.1 (1985): 147-169.

[8] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324

[9] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." Neural computation 18.7 (2006): 1527-1554.

[10] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." science 313.5786 (2006): 504-507.

[11] Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet classification with deep convolutional neural networks. In NIPS, pp. 1106–1114, 2012.

[12] Bottou, Léon. "Stochastic gradient descent tricks." Neural networks: Tricks of the trade. Springer Berlin Heidelberg, 2012. 421-436.

[13] Fukushima, Kunihiko. "Neocognitron: A hierarchical neural network capable of visual pattern recognition." Neural networks 1.2 (1988): 119- 130.

[14] V. Nair and G. Hinton, Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010.

[15] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "NNs Archtectures review," pp. 1–31, 2017.

[16] S. Hochreiter and J. Schmidhuber, \Long short term memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735

[17] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, \Show and tell: A neural image caption generator," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 260-269, 1967.

[18] W. Gerstner, "Spiking neurons", Pulsed Neural Networks, 2001.

[19] Rodrigo Carrasco-Davis, Guillermo Cabrera-Vives, Francisco Förster, Pablo A. Estevez, Pablo Huijse, Pavlos Protopapas, Ignacio Reyes, Jorge Martnez-Palomeraz, and Cristbal Donoso. "Deep Learning for Image Sequence Classification of Astronomical Events", arXiv:1807.03869

[20] Multiband Galaxy Morphologies for CLASH: A Convolutional Neural Network Transferred from CANDELS, M. Pérez-Carrasco1, G. Cabrera-Vives, M. Martinez-Marin, P. Cerulo, R. Demarco, P. Protopapas, J. Godoy1, and M. Huertas-Company, August 23 2019. Publications of the Astronomical Society of the Pacific, Volume 131, Number 1004

[21] Unsupervised Classification of Galaxies. I. Independent Component Analysis Feature Selection, Tanuka Chattopadhyay, Didier Fraix-Burnet, and Saptarshi Mondal, September 11th, 2019. The Astronomical Society of the Pacific. Publications of the Astronomical Society of the Pacific, Volume 131, Number 1004

[22] G. Segal, D. Parkinson, R. P. Norris, and J. Swan, “Identifying Complex Sources in Large Astronomical Data Sets Using a Coarse-grained Complexity Measure,” Publications of the Astronomical Society of the Pacific, vol. 131, no. 1004, p. 108007, 2019.

[23] L. Hendriks and C. Aerts, “Deep Learning Applied to the Asteroseismic Modeling of Stars with Coherent Oscillation Modes,” Publications of the Astronomical Society of the Pacific, vol. 131, no. 1004, p. 108001, 2019.

[24] S. Zucker and R. Giryes, “Shallow Transits—Deep Learning. I. Feasibility Study of Deep Learning to Detect Periodic Transits of Exoplanets,” The Astronomical Journal, vol. 155, no. 4, 2018.

[25] C. J. Shallue and A. Vanderburg, “Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90,” The Astronomical Journal, vol. 155, no. 2, p. 94, 2018.

[26] N. Krachmalnicoff and M. Tomasi, “Convolutional neural networks on the HEALPix sphere: a pixel-based algorithm and its application to CMB data analysis,” Astronomy & Astrophysics, vol. 628, p. A129, 2019.

[27] G. Longo, E. Merényi, and P. Tiňo, "Foreword to the Focus Issue on Machine Intelligence in Astronomy and Astrophysics," Publications of the Astronomical Society of the Pacific, vol. 131, no. 1004, p. 100101, 2019.

[28] R. Vilalta, K. D. Gupta, D. Boumber, and M. M. Meskhi, "A General Approach to Domain Adaptation with Applications in Astronomy," *Publications of the Astronomical Society of the Pacific*, vol. 131, no. 1004, p. 108008, 2019.

[29] [1] A. D'Isanto and K. L. Polsterer, "Photometric redshift estimation via deep learning Generalized and pre-classification-less, image based, fully probabilistic redshifts," Astronomy and Astrophysics, vol. 609, 2018.

[30] [1] A. D'Isanto and K. L. Polsterer, "Photometric redshift estimation via deep learning Generalized and pre-classification-less, image based, fully probabilistic redshifts," *Astronomy and Astrophysics*, vol. 609, 2018.

[31] Sloan Digital Sky Survey, https://www.sdss.org/, accessed: 11/12/2019

[32] M. delli Veneri, S. Cavuoti, M. Brescia, G. Riccio, and G. Longo, "Stellar formation rates in galaxies using Machine Learning models," pp. 1–6, 2018.

[33] E. E. O. Ishida, "Machine learning and the future of supernova cosmology," *Nature Astronomy*, vol. 3, no. 8, pp. 680–682, 2019.

[34] D. Baron, D. Poznanski, D. Watson, Y. Yao, N. L. J. Cox, and J. X. Prochaska, "Using machine learning to classify the diffuse interstellar bands," *Monthly Notices of the Royal Astronomical Society*, vol. 451, no. 1, pp. 332–352, 2015.

[35] [1] A. Hocking, J. E. Geach, Y. Sun, and N. Davey, "An automatic taxonomy of galaxy morphology using unsupervised machine learning," *Monthly Notices of the Royal Astronomical Society*, vol. 473, no. 1, pp. 1108–1129, 2018.

[36] S. D. Kügler and N. Gianniotis, "Modelling multimodal photometric redshift regression with noisy observations," vol. 7, no. November, p. 7, 2016.

[37] D. Baron and D. Poznanski, "The weirdest SDSS galaxies: Results from an outlier detection algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 465, no. 4, pp. 4530–4555, 2017.