Performance of Random Forest Machine Learning Algorithms in Binary Supernovae Classification

Jonathan Markel\textsuperscript{1,2}, Amanda J. Bayless\textsuperscript{2,3}

ABSTRACT

We consider random forest machine learning algorithms applied to the classification of Type Ia and core-collapse supernovae (CCSNe) by full light curves. We also quantitatively show the potential of early-epoch classification using the same machine learning techniques. The algorithm uses the shape and magnitude of the light curve peak to determine the classification. This an initial study to essentially determine SN type with as few data points as possible. New all-sky surveys will discover new transients at a rapid rate and decisions will have to be made on very little data which transients to follow-up. Here we present an initial method where as few as five data points near peak, we can identify whether it is a Type Ia or CCSNe with better than 80% accuracy. Furthermore, we provide an introduction into the use of machine learning for classification of astrophysical transients, including potential pitfalls, best practices, and areas in need of future research.

Subject headings: supernovae: general – methods: data analysis

1. Introduction

Supernovae (SNe) are the transient result of catastrophic star explosions. Astronomers discover several hundred SNe every year, and each one can provide valuable scientific insight into massive star composition and the environments in which they explode. The upcoming Large Synoptic Survey Telescope (LSST) will observe the sky in six filters (ugrizy) and is expected to detect 10 million SNe over its ten-year survey, or over 2000 each night.

\textsuperscript{1}The University of Texas at Austin, Department of Aerospace Engineering and Engineering Mechanics, 110 Inner Campus Drive, Austin, TX 78712, USA

\textsuperscript{2}Southwest Research Institute, Space Science and Engineering 6220 Culebra Rd. San Antonio, TX 78238, USA

\textsuperscript{3}The University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX 78249, USA
This will be a revolution in the number of SNe discovered. The most prolific SNe discovery surveys to date are the All-Sky Automated Survey for SuperNovae, which has discovered nearly 1000 bright SNe since 2014 (Holoien et al. 2017) and the Zwicky Transient Facility, which finds 3-4 SNe nightly and more than 1000 in the past year. The massive amount of data from the current surveys and the upcoming LSST is a hurdle for astronomers who seek to prioritize follow-up observational resources. This is a proof-of-concept study to define the differences in SNe using only observations, and to test early epoch performance of random forest machine learning algorithms for photometric classification. With the addition of LSST and follow up observations, the increase in the input of the number of light curves, particularly at early epochs, will allow a further refinement of the machine learning algorithm by type.

1.1. Supernovae Classification

Despite a wide-ranging spectrum and light curve-based taxonomy, most SNe are believed to be either thermonuclear explosions of white dwarf stars (Type Ia SNe) or the collapse of a star’s iron core. Core-collapse results in Type II SNe and is also believed to be the origin of Type Ib/c SNe. SNe are classified by their optical spectra into two types: Type II SNe exhibit hydrogen in their spectra and Type I SNe do not. Within Type I SNe, there are several subtypes: Type Ia SNe show strong Si II spectra lines; Type Ib SNe exhibit He I lines; Type Ic SNe exhibit neither Si II or He I lines. Type Ib/c are likely derived from striped-envelope stars (Filippenko 1997). Furthermore, subtypes within Type II SNe may depend on either spectra and photometric light curves. Type IIn SNe are so named because they exhibit narrow Balmer emission spectrum lines (Schlegel 1990) in early-epoch observations. Type IIb SNe display He signatures in later spectral observations. Type II-L and Type II-P are classified from their light curve characteristics (Filippenko 2005). After maximum brightness, Type II-L SNe decline linearly, and Type II-P SNe plateau for an extended period (a complete review of SNe subtypes can be found in Filippenko (1997)).

We need to observe SNe at as early of an epoch as possible. By observing at early epochs, we can better understand the evolution of the shock interactions. The shock dynamics shows valuable information about the physical properties of the progenitor stellar environment and in massive stars any mass loss material the star ejected into the surrounding adding to the interstellar medium (Ro & Matzner 2013). The shock interactions in all types of SNe occur at early times and it is difficult to reconstruct in models with only late time

1https://www.ztf.caltech.edu/news/zwicky-transient-facility-nabs-several-supernovae-a-night
In CCSNe the shock interactions are quite bright in the UV at early times (Sapir & Waxman 2017). Pritchard et al. (2014) also demonstrated that the bolometric luminosity in CCSNe is underestimated by more than 50% if the early-time UV observations are not included.

### 1.2. Machine Learning Background

A literature review revealed three key findings that guided our study:

1. Machine learning algorithms are very well suited to analyzing the large amounts of astronomical data expected from the large surveys and missions (Borne 2008; Karpinka et al. 2013; Kessler et al. 2010b; Lochner et al. 2016; Möller et al. 2016; Newling et al. 2011; Richards et al. 2011).

2. Large data sets of synthetically generated SNe light curves or spectra are necessary to train machine learning classifiers (Kessler et al. 2010a; Lochner et al. 2016; Möller et al. 2016; Newling et al. 2011).

3. The majority of research into SNe classification considers photometric observations from the beginning to the end of the explosion, and very few classify SNe during the first several observations (Kessler et al. 2010b).

Finding 2 guided the selection of data used to train the classification algorithm. Synthetic data is necessary as training requires 100s to 1000s of well sampled data sets. While 1000s of SNe have been discovered, there is not enough follow up data for each type to use observations as the training set. Additionally, SNe photometric and spectral observations are obtained with a "best-effort" approach to observe as many as possible, often without any a priori information. Consequently, many rare types of SNe (e.g. Type IInS) are missed and there is a bias towards observing bright SNe. Thus, there is not necessarily a representative sample of SNe types in spectroscopically confirmed observations, which would result in a biased classifier.

Finding 3 was especially significant to us. The earlier a rare SNe can be identified, the greater the chance that it can be measured with spectroscopy during its peak magnitude. Spectroscopy during peak can be very useful for determining the composition of a supernova, and more observations enables the refinement of theoretical models. The problem, then, is to identify SNe types as early as possible, using only light curves (as spectra will likely be unavailable), with a few data points as possible.
2. Methodology

In this section we discuss the classification algorithm used, including inputs (features), performance metrics, and implementation. We apply the same algorithm data in 2 different cases: classification using all the available light curve data, and using only the first few measurements (early-epoch classification). Additionally, we discuss the photometric light curve data used to classify SNe, including methods for parsing and potential biases of simulated SNe data.

2.1. Data

Several existing transient classification research papers utilized the 2010 Supernovae Classification Challenge (SNPhotCC) data set for testing (Karpenka et al. 2013; Newling et al. 2011; Revsbech et al. 2018). The SNPhotCC has since been followed up by the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC) (The PLAsTiCC team et al. 2018), the official findings of which are currently in progress. The SNPhotCC data consists of 21,318 simulated light curves and is the most extensive and available representative datasets of SNe, making it especially useful for testing machine learning classification algorithms. It contains simulated Type Ia, Ib/c, and II SNe observations in the $griz$ filters based on the Dark Energy Survey (DES). Non-type Ia SNe simulations are based on spectrographically confirmed light curves donated by the Carnegie Supernova Project, the Supernova Legacy Survey and the Sloan Digital Sky Survey. The contribution of non-Ia data from these surveys likely undersamples the variety of SNe that will be observed by LSST, but will be a useful step away from the overly-simplistic studies that have relied on a handful of non-Ia templates (Kessler et al. 2010a).

We also had access to the Supernova Analysis Application (SNAP) database (Bayless et al. 2017), which currently contains several hundred observed SNe from space and ground observations. However, the SNPhotCC dataset has several advantages over SNAP for this area of research. First, the SNAP database contains SNe from multiple surveys, each with slightly different sets of filters, and this non-homogeneity presents special challenges when comparing the same SNe metrics across different sky surveys. More importantly, the literature review and large number of observations expected from LSST led us to conclude that machine learning algorithms would be best suited for the scale of autonomous SNe classification task, and machine learning algorithms require large amounts of data to adequately train. While the

https://snap.space.swri.edu/snap/>
number of SNe in SNAP is notable, we opted to use a dataset of synthetically generated SNe to train the classifier on a greater variety of types.

It is also worth noting that a new synthetic dataset generated specifically for LSST was due to be released in the fall of 2018 (The PLAsTiCC team et al. 2018). Unfortunately, a majority of our research took place in summer before this data was released and we were not able to utilize it here. Nonetheless, preparatory materials for PLAsTiCC recommend utilizing the SNPhotCC dataset for building a software framework as well. A preprint draft outlining the models and simulations used in the PLAsTiCC dataset is available as of March 2019 (Kessler et al. 2019).

2.1.1. Data Cleaning

Raw data of astronomical observations often come with noise and other factors that must be accounted for. Additionally, an astronomical transient may not always be observed in every filter. As a result, there may be a filter without observations. While some algorithms like XGBoost (Chen & Guestrin 2016) are capable of handling missing values, they are less robust than random forests because they train each decision tree sequentially and based on previous tree performance. In contrast, random forests train each tree individually, making them less prone to overfitting. The scikit-learn implementation of random forests does not handle missing values in so missing data must be imputed without biasing the model. After calculating values for each SNe feature, we impute missing values in the training set with the median value for each column of non-zero observations per Breiman & Cutler (2004).

2.2. Random Forest Classification

Boosted decision trees and random forest classifiers have been shown to be one of the most robust algorithms for photometric classification when compared to naive Bayes, k-nearest neighbors, support vector machines, and convolutional neural networks (Lochner et al. 2016). Decision trees map input features to output classes based on specific rules. The model is then trained by recursively selecting features and boundaries which increase the overall information gain. In other words, decision trees find boundaries which minimize the entropy of the working set. A single decision tree has its drawbacks – it can be prone to over-fitting and sensitive to small changes in the input data. However, these effects can be mitigated by combining multiple decision trees into an ensemble, where the final classification is determined by a majority vote of the many trees contained within the model. A key benefit of
ensemble decision tree algorithms like random forests is that the confidence of any classification can be directly interpreted from the proportions of decisions of individual trees. We utilized the scikit-learn random forest classifier, an ensemble method that combines many decision trees.

2.2.1. Model Parameters

The number or estimators, or trees, in a random forest classifier can impact the models performance. Random forests cannot be overfit (Breiman 2001), but after a certain number of trees, increasing the number of estimators has a detrimental effect on computational runtime and does not improve performance. However, more estimators do increase the stability of the model. We determine the optimal number of trees for our classifier by calculating the out-of-bag score (equal to $1 - [\text{out-of-bag error}]$) for models ranging from 1 to 128 estimators. After analyzing performance trade-offs, we selected 40 trees.

2.3. Feature Selection

We consider two techniques for feature extraction. First, we considered a general parametrization fitted with the following form, which we refer to as the “Bazin Form”:

$$f^k(t) = A^k \frac{e^{-\frac{(t-t_0)}{\tau_{\text{fall}}}}}{1 + e^{-\frac{(t-t_0)}{\tau_{\text{rise}}}}} + c^k$$

(1)

where $t_{\text{fall}}$ and $t_{\text{rise}}$ are the times before and after peak magnitude, $c$ is a constant, $A$ normalizes the signal, and $t_0$ relates the rise and fall times to the time of peak magnitude (Bazin et al. 2009; Dai et al. 2018). The five Bazin fit parameters do not represent any physical characteristics but are sufficiently general to fit all SNe types. This model is used in other photometric classification literature to reduce the dimensionality of light curve data (Karpenka et al. 2013; Möller et al. 2016; Dai 2017; Ishida et al. 2019).

In Figures 1 and 2, we show examples of both acceptable and poorly fitted data. In Figure 1, the model does decently representing the underlying data points across the $g$, $r$, and $i$ bands. However, the fit does not perfectly converge in the $z$ band.

In Figure 2, none of the fits adequately represent the underlying data. Additionally,
Fig. 1.— An example of an acceptable fit using the Bazin model.

this SN lacks observations in the $g$ band. This does not have an effect on fitting the other bands, as the underlying model fits each filter independently.

Secondly, we describe light curves using statistics that measure the variability and characteristics of a light curve. This technique consists of using six features to describe the curve in each filter, namely maximum magnitude, kurtosis, skew, median absolute deviation, the Shapiro-Wilk statistic (Shapiro & Wilk 1965), and the coefficient of variation (Table 1). Each feature was normalized by removing the mean and scaling to unit variance. We use these parameters to implicitly reflect light curve characteristics which are known to be indicators of physical parameters of core-collapse SNe (CCSNe), such as the explosion energy, envelope composition, or mass.

2.4. Early-Epoch Classification

The SNPhotCC gave participants the extra challenge to classify SNe using only the first six observations in any filter, however, it was not attempted by any participants due to time constraints and increased interest in the full light curve challenge (Kessler et al. 2010b). Given our interest in early-epoch classification, we use the challenge constraint of six points as a baseline. We add an additional constraint for our early-epoch analysis – being able to determine type before the SNe reaches its greatest apparent magnitude. However, some simulated SNe in the SNPhotCC have fewer than 6 observations before the peak magnitude
Fig. 2.— An example of an unacceptable fit using the Bazin model.

due to the edge effects of simulating Dark Energy Survey data. The SNPhotCC also considers the time constraints of SN surveying, and not all light curves start at explosion.

In our early-epoch analysis of the SNPhotCC data, we identify the peak magnitude in each light curve and applied a quality cut wherein any SNe with fewer than five data points before the peak in two or more filter bands are removed. This rule results in the removal of 5,609 SNe, or 26.3% of the data set. For the approximately 15,700 remaining SNe, all post-peak observational data was discarded. Each filter band was then fit to a quadratic polynomial, and filter bands with fewer than three points were fit to a line.

Many of the statistical features used to analyze full light curves are dependent on a full distribution with an identifiable peak. With our early epoch data set, we cannot assume knowledge of a defined peak magnitude. A reduced set of statistical features and one new feature were used, and can be found in Table 2. The early-epoch classification portion of this research includes two cases – one in which the statistical features are calculated from discrete observational data, and another where the data is fitted to a third-order polynomial.

2.5. Performance Metrics

To measure the performance of our model, we use standard performance metrics such as precision, recall, and the F1 score \citep{vanRijsbergen1979}, which is the harmonic mean
| Value            | Formula                                                                 | Feature Description |
|------------------|-------------------------------------------------------------------------|---------------------|
| Maximum Peak     | \( \text{max}(A) \)                                                    | Peak magnitude value |
| Kurtosis         | \( n \frac{\sum_{i=1}^{n} (X_i - \text{avg})^4}{(\sum_{i=1}^{n} (X_i - \text{avg})^2)^2} \) | "Sharpness" of distribution |
| Skew             | \( \frac{\bar{X} - Mo}{\sigma} \)                                     | Distribution asymmetry measure |
| Medium Absolute Deviation | \( \text{median}(X_i - \text{median}(X)) \)                      | Measure of population variability |
| Coefficient of Variation (CV) | \( \frac{\sigma}{\bar{X}} \)                                   | Normalized variability to mean relation |
| Shapiro-Wilk Statistic | \( \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \) | Measure of distribution normality |

Table 1: Full light curve statistical features used in classification.

of precision and recall. Precision \((\text{Precision} = \frac{TP}{TP+FP})\) is the number of true positives (TP) divided by the combined true positives and false positives (FP), and it can be thought of as a measure of the classifier's exactness. The recall \((\text{Recall} = \frac{TP}{TP+FN})\) is the number of true positives divided by the combined true positives and false negatives (FN). Recall can be thought of as the classifier's ability to classify the full set of positives. Additionally, positive class recall is referred to as the sensitivity, and negative class recall is referred to as specificity. The F\(_1\) score is the weighted average of recall and precision \((\text{Yang & Liu 1999})\), and is calculated as \(F_1(r,p) = \frac{2rp}{r+p}\). The F\(_1\) score is more useful than standard metrics when the classes are unevenly distributed, as they are in the SNPhotCC data.

Finally, we use the out-of-bag error estimate \((\text{Breiman 1996b})\). Individual trees in a random forest are constructed using subsets of the total data. For each tree, testing performance on the unused subset produces an internal error estimate. Combining these
Table 2: Early epoch classification features.

| Value                     | Formula                                                                 | Feature Description                           |
|---------------------------|-------------------------------------------------------------------------|-----------------------------------------------|
| Average Rate of Change    | $\frac{\sum_{i=1}^{n}(x_{i+1}-x_i)}{n}$                              | Slope of the rise to peak                     |
| Coefficient of Variation  | $\frac{\sigma}{\bar{X}}$                                              | Normalized variability to mean relation       |
| (CV)                      |                                                                         |                                               |
| Skew                      | $\frac{\bar{X}-M_0}{\sigma}$                                          | Distribution asymmetry measure                |
| Kurtosis                  | $n \frac{\sum_{i=1}^{n}(X_i-X_{avg})^4}{(\sum_{i=1}^{n}(X_i-X_{avg})^2)^2}$ | "Sharpness" of the distribution              |

3. Results

One advantage of random forests is the ease with which the relative importance of any given feature for correctly classifying a SNe can be determined. The implementation of random forests in scikit-learn includes a function for determining a features importance. However, Strobl et al. (2007) shows that these feature importances can be biased. Instead of using the built in feature importance function, we determine drop column importance for each feature (Parr et al., 2018). The drop column importance of a feature is calculated by measuring the decrease in classification performance when a given feature is removed from the model. If the model performance decreases significantly without the feature, it is assigned a higher importance. Additionally, we introduce a column of randomly generated numbers to help determine whether any given feature is contributing to classification performance. By comparing feature importance to a set of randomly generated numbers as a baseline, we can identify useless and even detrimental features to be those which are less important the
random number set. To visualize the importance of features we normalize their importance relative to the importance of the randomly generated number set. The feature importances are normalized according to Equation 2. $X'$ is the drop column importance value calculated with the code supplied in Parr et al. (2018), $X_{\text{random}}$ is the importance of the set of randomly generated values, and $\sigma_X$ is the standard deviation of the set of feature importances.

$$X' = \frac{X - X_{\text{random}}}{\sigma_X}$$ (2)

We compare the performance of random forest classifiers when applied to full light curves and pre-peak light curves. Additionally, we classify SNe without interpolating data over a fitted model to provide a baseline performance. In all, we consider four variations:

1. Full light curves, where 6 features are calculated from discrete data points in each filter.
2. Full light curves, where 6 features are calculated from interpolated light curve data.
3. Pre-peak light curves, where 4 features are calculated from discrete data points in each filter.
4. Pre-peak light curves, where 4 features are calculated from interpolated light curve data.

In both early-epoch and full light curve cases, it can be seen that interpolation of data – even when not perfectly fit to a model – improves classification performance.

3.1. Full Light Curve Classification

Comparing classification performance between raw data (Table 3) and fitted data (Table 4), it can be seen that interpolating data is a highly-effective way of increasing classifier capability. Specifically, we see an improvement in core-collapse SNe recall, and Type Ia precision. In other words, with fitted data, our classifier correctly typed a greater portion of core-collapse SNe and was more capable of correctly classifying Type Ia supernovae.

We normalized the importance of individual features relative to a “feature” of randomly generated numbers. By fitting our data to a model and interpolating, we greatly reduce the overall randomness of features used for classification (compare Figures 3 and 4). Certain features are particularly important for correctly classifying a SN from light curve data, like
|                | Precision | Recall | F₁ Score |
|----------------|-----------|--------|----------|
| Type Ia        | 0.71      | 0.49   | 0.58     |
| Core Collapse  | 0.85      | 0.93   | 0.89     |
| Average/Total  | 0.81      | 0.82   | 0.81     |

Table 3: Classification performance for \textit{unfitted}, full light curves.

|                | Precision | Recall | F₁ Score |
|----------------|-----------|--------|----------|
| Type Ia        | 0.77      | 0.49   | 0.60     |
| Core Collapse  | 0.85      | 0.95   | 0.90     |
| Average/Total  | 0.83      | 0.84   | 0.82     |

Table 4: Classification performance for \textit{fitted}, full light curves.

the peak magnitude value. While the magnitude in the $z$ and $i$ filters were more important for fitting unfitted data, the $g$ and $i$ filters were more important for classifying SN with fitted light curves. From the importance of the peak magnitude and variation across filters, we infer that the peak magnitude is key to classifying SN no matter which observing filter is used.

Furthermore, when classifying actual SNe, one should remove any feature which is less important than the random feature. They are irrelevant at best and potentially harmful to the result.

![Feature Importances (Unfitted, Full Light Curve Classification)](image)

Fig. 3.— Drop column importance of features in a classifier trained on unfitted data.
3.2. Early-Epoch Classification

We see the same improvement in performance for early-epoch classification after fitting our data to a third-order polynomial. Comparing Tables 5 and 6 shows an increase in core-collapse precision, while recall remains steady. This is in contrast to the full light curve classification task.

|                | Precision | Recall | F_1 Score |
|----------------|-----------|--------|-----------|
| Type Ia        | 0.61      | 0.28   | 0.38      |
| Core Collapse  | 0.79      | 0.94   | 0.86      |
| Average/Total  | 0.74      | 0.77   | 0.74      |

Table 5: Classification performance for *unfitted* early-epoch light curves.

|                | Precision | Recall | F_1 Score |
|----------------|-----------|--------|-----------|
| Type Ia        | 0.72      | 0.45   | 0.55      |
| Core Collapse  | 0.84      | 0.94   | 0.89      |
| Average/Total  | 0.81      | 0.82   | 0.80      |

Table 6: Classification performance for *fitted* early-epoch light curves.

Figures 5 and 6 confirm that the rate of rise to peak can be a useful indicator for early-epoch SN classification. In the absence of a defined peak, the skew of the data was highly important for both the unfitted and fitted data. We also see a decrease in feature randomness after fitting and interpolating data, similar to the full light curve case.
In this paper we discuss our process of creating a machine learning classifier for SNe. We successfully identified key areas for continued research, potential pitfalls for researchers getting started in astrophysical classification, and established a software framework for our future work to build upon. This is the start of a more targeted approach to identify SN types, particularly rare types, with limited information. This is difficult but necessary as there will be few light curve data points observed by LSST in early epochs.

4. Discussion
4.1. Necessity of Interpolating Light Curves

The greatest challenge we faced in our research was fitting unclassified SNe models to an underlying model.

The model proposed by Bazin et al. (2009) is general enough to fit most SNe, which is especially beneficial when the target SNe's type is unknown. However, a significant number of light curves fit with this model either did not converge or were not fit well enough to describe the underlying observations (see Figure 2). As a result, SNe which did not fit the Bazin model had to be removed or fit with another model. Poor fits may be partially mitigated by iteratively refitting with different parameters, as applied in Dai et al. (2018).

One area of future research is to combine the iterative approach applied by Dai et al. (2018) with type-specific models. For example, once a SNe is fit to the general model and an initial guess at type is found, the SNe could be fit to a number of type-specific models, depending on classification uncertainty, in an attempt to improve classification. Once the best fit is found, the data could be interpolated, reclassified, and used to improve future predictions. This approach would also have the advantage of providing an automated, statistical means of validating type-specific models with each spectroscopically confirmed SNe.

4.2. Feature Importance

From the cases studied, we find that not all features are created equal in random forest classification. Clearly, the peak magnitude is key to classification of full light curves, as the peak magnitude value was a highly important feature (see Figures 3 and 4). Future researchers should consider the value in selecting unique features for each filter band, dependent upon the data being analyzed and the performance of their own model. It is highly recommended to remove any feature which is outperformed by a column of randomly generated numbers.

Furthermore, our study of feature importance underscores the need for fitting data to a model. For example, in full light curve classification, comparing Figures 3 and 4 shows how even inconsistent fitting drastically reduces the randomness in a model. The same is true for the early-epoch case, when most SNe were fit to a line or spline.
5. Conclusion

As computational power and the scale of astronomical surveys continue to improve, machine learning algorithms – and especially random forest classifiers – can be a robust and useful tool for understanding our universe. When used for transient classification, machine learning can help to quickly identify events worth closer examination. As our ability to identify SNe improves through the use of new technologies and methods, so will our understanding of the physical mechanisms behind these fundamental astronomical events.

This work is supported by the NASA Astrophysical Data Analysis Program (NNH15ZDA001N-ADAP).

REFERENCES

Bayless, A. J., Fryer, C. L., Wollaeger, R., et al. 2017, The Astrophysical Journal, 846, 101

Bazin, G., Palanque-Delabrouille, N., Rich, J., Ruhlmann-Kleider, V., & Aubourg, E., et. a.
2009, A&A, 499, 653

Borne, K. 2008, Astronomische Nachrichten: Astronomical Notes, 329, 255

Breiman, L. 1996a, Machine learning, 24, 123

—. 1996b, Out-of-bag estimation

—. 2001, Machine learning, 45, 5

Breiman, L., & Cutler, A. 2004

Chen, T., & Guestrin, C. 2016, in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM, 785–794

Dai, M. 2017, Cosmology with Type Ia Supernovae: Robust Techniques for Data Analysis and Photometric Classification

Dai, M., Kuhlmann, S., Wang, Y., & Kovacs, E. 2018, MNRAS, 477, 4142

Filippenko, A. V. 1997, Annual Review of Astronomy and Astrophysics, 35, 309

Filippenko, A. V. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 342, 1604-2004: Supernovae as Cosmological Lighthouses, ed. M. Turatto, S. Benetti, L. Zampieri, & W. Shea, 87
Holoien, T.-S., Brown, J., Stanek, K., et al. 2017, Monthly Notices of the Royal Astronomical Society, 471, 4966

Ishida, E. E. O., Beck, R., González-Gaitán, S., et al. 2019, MNRAS, 483, 2

Karpenka, N. V., Feroz, F., & Hobson, M. P. 2013, MNRAS, 429, 1278

Kessler, R., Conley, A., Jha, S., & Kuhlmann, S. 2010a, arXiv preprint arXiv:1001.5210

Kessler, R., Bassett, B., Belov, P., et al. 2010b, Publications of the Astronomical Society of the Pacific, 122, 1415

Kessler, R., Narayan, G., Avelino, A., et al. 2019, arXiv preprint arXiv:1903.11756

Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, The Astrophysical Journal Supplement Series, 225, 31

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201

Möller, A., Ruhlmann-Kleider, V., Leloup, C., et al. 2016, J. Cosmology Astropart. Phys., 12, 008

Newling, J., Varughese, M., Bassett, B., et al. 2011, MNRAS, 414, 1987

Parr, T., Turgutlu, K., Csiszar, C., & Howard, J. 2018, Beware Default Random Forest Importances

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., & Thirion, B., e. a. 2011, Journal of Machine Learning Research, 12, 2825

Pritchard, T., Roming, P., Brown, P. J., Bayless, A. J., & Frey, L. H. 2014, The Astrophysical Journal, 787, 157

Revsbech, E. A., Trotta, R., & van Dyk, D. A. 2018, MNRAS, 473, 3969

Richards, J. W., Homrighausen, D., Freeman, P. E., Schafer, C. M., & Poznanski, D. 2011, Monthly Notices of the Royal Astronomical Society, 419, 1121

Ro, S., & Matzner, C. D. 2013, The Astrophysical Journal, 773, 79

Sapir, N., & Waxman, E. 2017, The Astrophysical Journal, 838, 130

Schlegel, E. M. 1990, MNRAS, 244, 269
Shapiro, S. S., & Wilk, M. B. 1965, Biometrika, 52, 591

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. 2007, BMC bioinformatics, 8, 25

The PLAsTiCC team, Allam, Tarek, J., Bahmanyar, A., et al. 2018, arXiv e-prints, arXiv:1810.00001

van Rijsbergen, C. J. 1979, Information Retrieval (Butterworth)

Yang, Y., & Liu, X. 1999, in Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’99 (New York, NY, USA: ACM), 42–49

This preprint was prepared with the AAS LaTeX macros v5.2.