Classification of White Dwarfs Observed by SDSS

By: Bo Xia, Kaylin Li, Roochi Shah, Yixuan Wu

Advisor: Peter Freeman

INTRODUCTION

The Sloan Digital Sky Survey (SDSS) has observed high-resolution spectra, in addition to brightness over five different bandpasses, for many objects known as white dwarfs (WDs). WDs are remnants of low-mass stars like the Sun, and their spectra are historically classified into a number of types. Our goal is to see if we can **identify WDs of spectral type DA given easily obtained, non-spectroscopic information**.

DATA

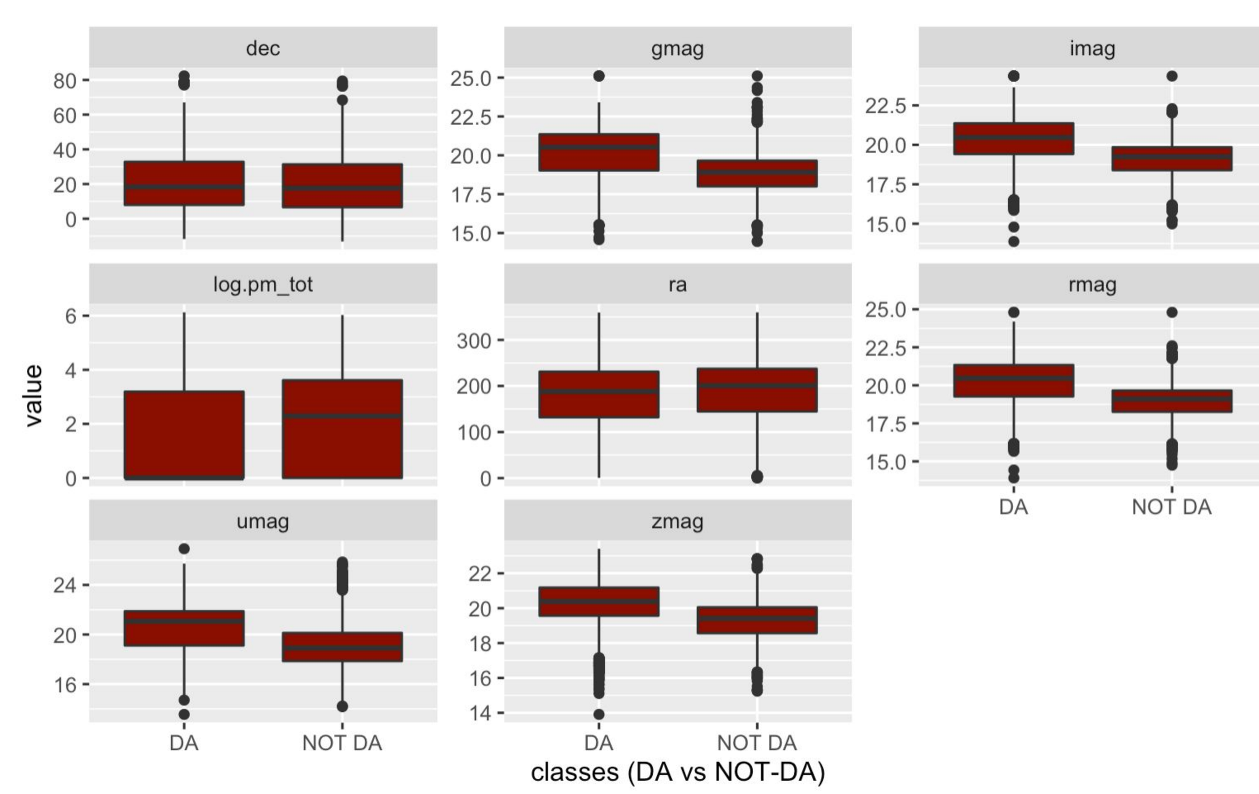
The dataset based on the catalog of Kepler et al. (2014) contains information for 9,112 white dwarfs labeled as being either DA (spectral type A) or NOT-DA. There are eight predictor variables:

```
       ra                dec               gmag            umag            rmag
 Min.   :  0.2986   Min.   :-13.149   Min.   :14.45   Min.   :13.56   Min.   :13.91
 1st Qu.:133.6356   1st Qu.:  7.737   1st Qu.:18.74   1st Qu.:18.80   1st Qu.:18.98
 Median :191.3272   Median : 18.182   Median :20.12   Median :20.70   Median :20.03
 Mean   :182.3488   Mean   : 20.203   Mean   :19.85   Mean   :20.23   Mean   :19.94
 3rd Qu.:232.2706   3rd Qu.: 32.473   3rd Qu.:21.21   3rd Qu.:21.76   3rd Qu.:21.18
 Max.   :359.9055   Max.   : 82.332   Max.   :25.11   Max.   :26.92   Max.   :24.80
     pm_tot           class                            imag            zmag
 Min.   :  0.00   DA    :7240                    Min.   :13.87   Min.   : 0.03
 1st Qu.:  0.00   NOT DA:1872                    1st Qu.:19.17   1st Qu.:19.34
 Median :  0.00                                  Median :20.07   Median :20.16
 Mean   : 17.83                                  Mean   :20.04   Mean   :20.07
 3rd Qu.: 26.00                                  3rd Qu.:21.20   3rd Qu.:21.05
 Max.   :456.70                                  Max.   :24.37   Max.   :23.41
```

Variables	Description
ra, dec	celestial longitude and latitude
umag, gmag, rmag, imag, zmag	magnitude in each SDSS bandpass
pm_tot	yearly motion relative to background stars

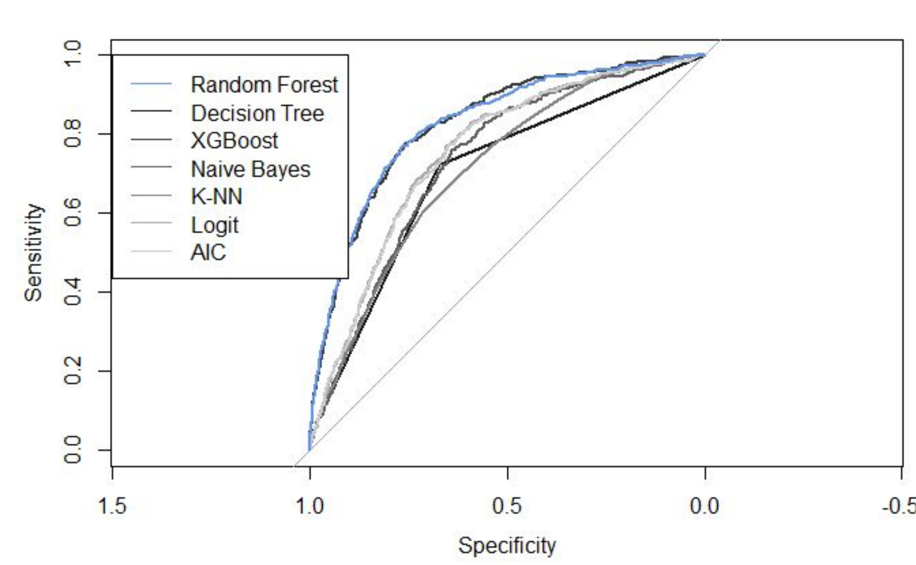
The response classes are imbalanced, with 79.5% DA and 20.5% NOT-DA. We remove one row with an outlier value for **`zmag`**. The summaries for the five magnitudes are similar. As for **`pm_tot`**, its values are 0 or >2, and to make the positive values less skew, we performed a logarithmic transformation.

The figure to the right shows that there are significant differences in the magnitude variables between the two classes, which might indicate possible associations. Meanwhile, **`ra`** and **`dec`** are relatively similar between the two classes.



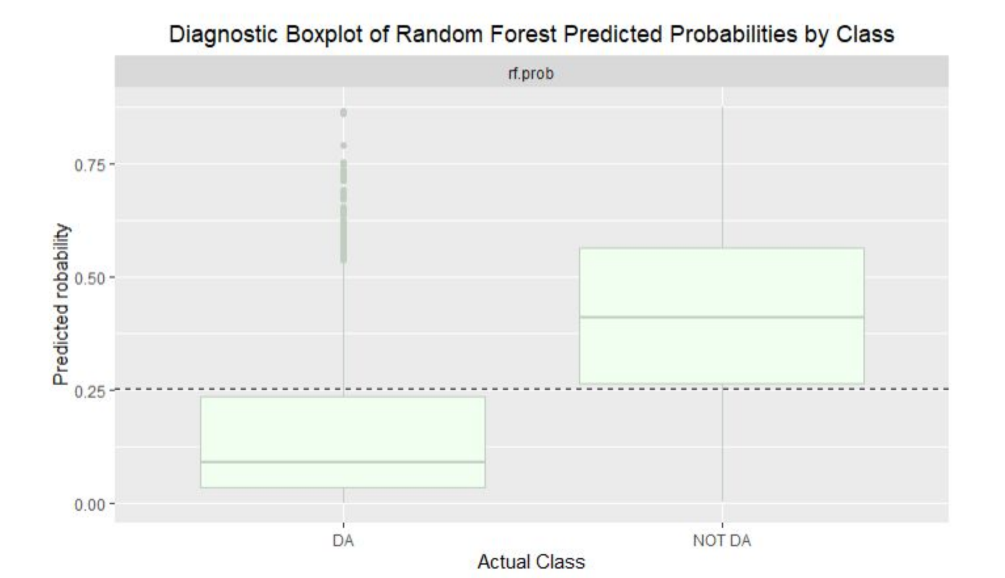
ANALYSIS

We split the dataset, retaining 75% for training and 25% for testing, and learn multiple classification models, as listed in the table below. We compute a receiver operating characteristics (ROC) curve for each model, which illustrates the tradeoff between making accurate predictions in each class. We select the model with the largest area under the ROC curve (AUC) and utilize Youden's J statistic to generate class predictions. In the figure at right, the estimated probability of being NOT-DA is shown for objects of each class. The dotted line corresponds to the optimized Youden's J value; WDs with probabilities below the line, for instance, are predicted to be of class DA.The metric used to evaluate classification success was AUC, or area under the ROC curve. Out of the models tested, Random Forest had the highest AUC of approximately 0.826.




```
         resp.test
rf.pred    DA NOT DA
  DA     1383    111
  NOT DA  419    365
```

Model	MCR	AUC
Logistic Regression	0.323	0.753
AIC	0.348	0.752
Decision Tree	0.317	0.702
Random Forest	0.233	0.826
XGBoost	0.237	0.825
Naive Bayes	0.336	0.732
K-NN	0.691	0.713



CONCLUSION

Given our dataset, we found that the best model for predicting whether a white dwarf was of spectral type A was a Random Forest model with an AUC of 0.826 and a misclassification rate of 0.233. Therefore, we can conclude that we can **determine the spectral type of a white dwarf given information about its brightness, location on the sky, and apparent movement on the sky.** A possible next step to improve prediction accuracy is collecting more data on white dwarfs that are not of spectral type A.

References:

Kepler, S. O., et al. (2014). New white dwarf stars in the Sloan Digital Sky Survey Data Release 10. *Monthly Notices of the Royal Astronomical Society*, 1-10. https://arxiv.org/pdf/1411.4149.pdf

Introduction: Freeman, P. E. 2021, online at https://github.com/pefreeman/36-290/blob/master/PROJECT DATASETS/WD CLASS/README.md