MK CLASSIFICATION OF SPECTRA USING AN AUTOMATED CLASSIFICATION ALGORITHM

OCTOBER 30, 2018

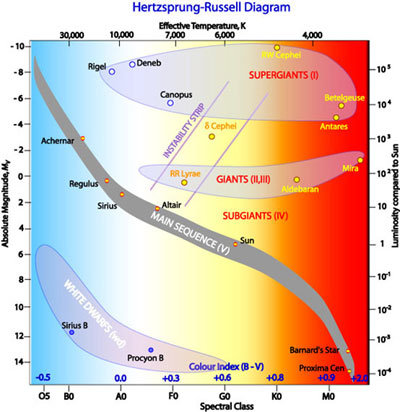
ROSIE FASULLO, DAVID WHELAN, MIKE HIGGS

ABSTRACT

This paper presents the work that has been done to develop an automated spectral classification algorithm. The algorithm describes spectra according to the Lorentzian probability density function that fits the wings of the hydrogen Balmer lines Hβ-H8. The algorithm then utilizes the random forest classifier from Sci-Kit Learn to create a classification model based on a selection of 96 confidently classified spectra from the data set at Austin College. Using k-fold cross validation to measure the average accuracy of the algorithm, the current model is correctly classifying data at a rate of 79%. The project will be continued in January of 2019 and continued through the spring semester of the 2018-2019 academic year. To increase the accuracy of the algorithm we are going perform hyperparameter tuning on the model itself, obtain more data from various sources and add a description of the helium metallic lines for the algorithm to use in classification.

1. INTRODUCTION

Stars are grouped according to two different criteria, spectral type and luminosity class. The spectral type corresponds to the relative color of the light emitted by a star. The color of a star is an indication of what the surface temperature of the star is; the bluer the color, the hotter the star (Gray and Corbally 2009). There are 7 different spectral types used by astronomers today and they are, from bluest to reddest, O, B, A, F, G, K, and M (Gray and Corbally 2009). For reference, our sun is a spectral type G star. Stars above this type, the K and M type stars, are referred to as late-type stars. Stars that are below this type, the O, B, A and F type stars, are referred to as early-type stars (Gray and Corbally 2009). Within each type there is a numeric subtype ranging from 0 to 9 with 0 corresponding to the bluest subtype and 9 corresponding to the reddest subtype (Gray and Corbally 2009). Our sun has a spectral type of G2, meaning it is on the hotter end of a mid-range star (Gray and Corbally 2009). Luminosity class corresponds to the brightness of the star, which is an indication of the size of the star (Gray and Corbally 2009). Stars range from dwarf stars with luminosity class V to supergiant stars with luminosity class Ia (Gray and Corbally 2009). Our sun is a luminosity class V star, which is the most common luminosity class of the stars astronomers have observed (Gray and Corbally 2009). The distribution of spectral types and luminosity classes can be seen in what is called a Hertzsprung-Russell (HR) Diagram in Figure 1 below. The current observing project at the Adams Observatory is the spectral classifications of O, B and A type stars, with the goal of better understanding their fundamental properties and types of chemical peculiarities.



*Figure 1: Hertzsprung-Russell Diagram. Effective temperature is listed on the top the graph with the corresponding spectral type along the bottom. The luminosity is shown on the left with the corresponding absolute magnitude of the star along the right. The higher the magnitude of brightness the dimmer the star (“Hertzsprung-Russel Diagram”).*

In addition to spectral types and luminosity classes, a notation of any spectral peculiarities is also included in spectral classifications. There are several different types of spectral peculiarities. One form of spectral peculiarity is full or partial emission in the hydrogen lines. Emission normally indicates there is a shell or disk around the star emitting light as the light from the star excited atoms in the disk (Gray and Corbally 2009). A chemically peculiar star shows unusual metal lines in its spectrum. Many chemically peculiar stars show unusually weak or unusually strong helium lines, indicating either an abundance or a lack of helium in the star (Gray and Corbally 2009). This can make these stars tricky to classify because helium lines are a critical part of classifying O and B type star spectra (Gray and Corbally 2009). There may also be unusually strong absorption of other metallic lines, indicating an unusually large abundance of that metal. Another common type of peculiarity is a binary star system. Depending on the physical relationship between the stars, the binary nature of the system may not be photometrically visible. Thus, the spectrum of the system may be the only way to see the spectral absorption features from both stars (Gray and Corbally 2009). A composite spectrum is distinguished from a normal spectrum by the presence of two spectra; although this can be quite difficult if the two stars are physically similar (Gray and Corbally 2009). Nebular spectra are characterized by the broadening and shortening of the metallic lines. What distinguishes nebular spectra from normal spectra is that although the depth and width of the lines are different than the standard, the strength (or total area of the line) is the same (Gray and Corbally 2009). A nebular spectrum indicates that the star is rotating at an unusually high rate and the widening of the lines is due to the Doppler effect (Gray and Corbally 2009).

There are several different systems used to classify stars. The system we employ is the MK classification system developed by W. W. Morgan and P.C. Keenan (Morgan and Kennan 1973). The MK system classifies stars by comparing spectrum of the star in question to a set of spectral standards and giving the star the label of whichever spectral standard it most resembles (Gray and Corbally 2009). This comparison is done using only spectra and does not admit the use of any extra information about the star such as photometric data (Gray and Corbally 2009). In this way, the MK system classifies stars according to their most natural groupings. The system does allow for the interpolation of spectral classifications if a star appears to lie between two spectra (Gray and Corbally 2009). The MK system allows spectral standards of three different types per classification: anchor points, primary standards, and secondary standards (Gray and Corbally 2009). An anchor point is a spectral standard that has not changed classification since development of the MK classification system began (Gray and Corbally 2009). Primary standards fill in the gaps of the anchor points and provide a reference for the best-known spectrum for each spectral classification (Gray and Corbally 2009). Secondary standards are the best-known stars that are accessible from both the northern and southern hemisphere (Gray and Corbally 2009).

The classification of stars is a key part of understanding the physical characteristics of stars. However, stellar classification is a time consuming and difficult task and as such is often neglected by astronomers. Therefore, an accurate and efficient spectral typing program would be hugely beneficial. What a team of astronomers would do over the course of several days, a reliable spectral typing program could do in a few hours. Unfortunately, although a few spectral classification algorithms have been created, the most reliable method is still classifying stars by eye. The MK classification system often relies heavily on the astronomer’s intuition. Although computers are good at many things, they are unable to make decisions based on intuition and simulating intuition is very complex. We hope to simulate this intuition using artificial intelligence.

There has been an attempt at an expert classification program by R. Gary and C. Corbally. They developed a program called *MKCLASS* that can accurately classify normal and chemically peculiar stars within 0.6 of a spectral type and 0.5 of a luminosity class (Gray and Corbally 2014). *MKCLASS* uses a weighted least squares comparison combined with a detailed comparison of the metallic line strengths to determine the spectral classification (Gray and Corbally 2014). *MKCLASS* is very good at classifying stars that are normal or have chemical peculiarities (Gray and Corbally 2014). However, there are some drawbacks to the *MKCLASS* program. Unfortunately, when a spectrum has a peculiarity that is not expected by the program, it gives up at classifying the spectrum without an explanation as to why the classification failed (Gray and Corbally 2014). In addition, there is no way to numerically measure the accuracy of the spectral classification of the program (Gray and Corbally 2014). The program provides a file that lists the quality of the classification that ranges from “Very poor” to “Very good”, but this refers to the signal to noise ratio of the spectrum, not the accuracy of the classification (Gray 2015). While *MKCLASS* is a strong program there is certainly room for improvement.

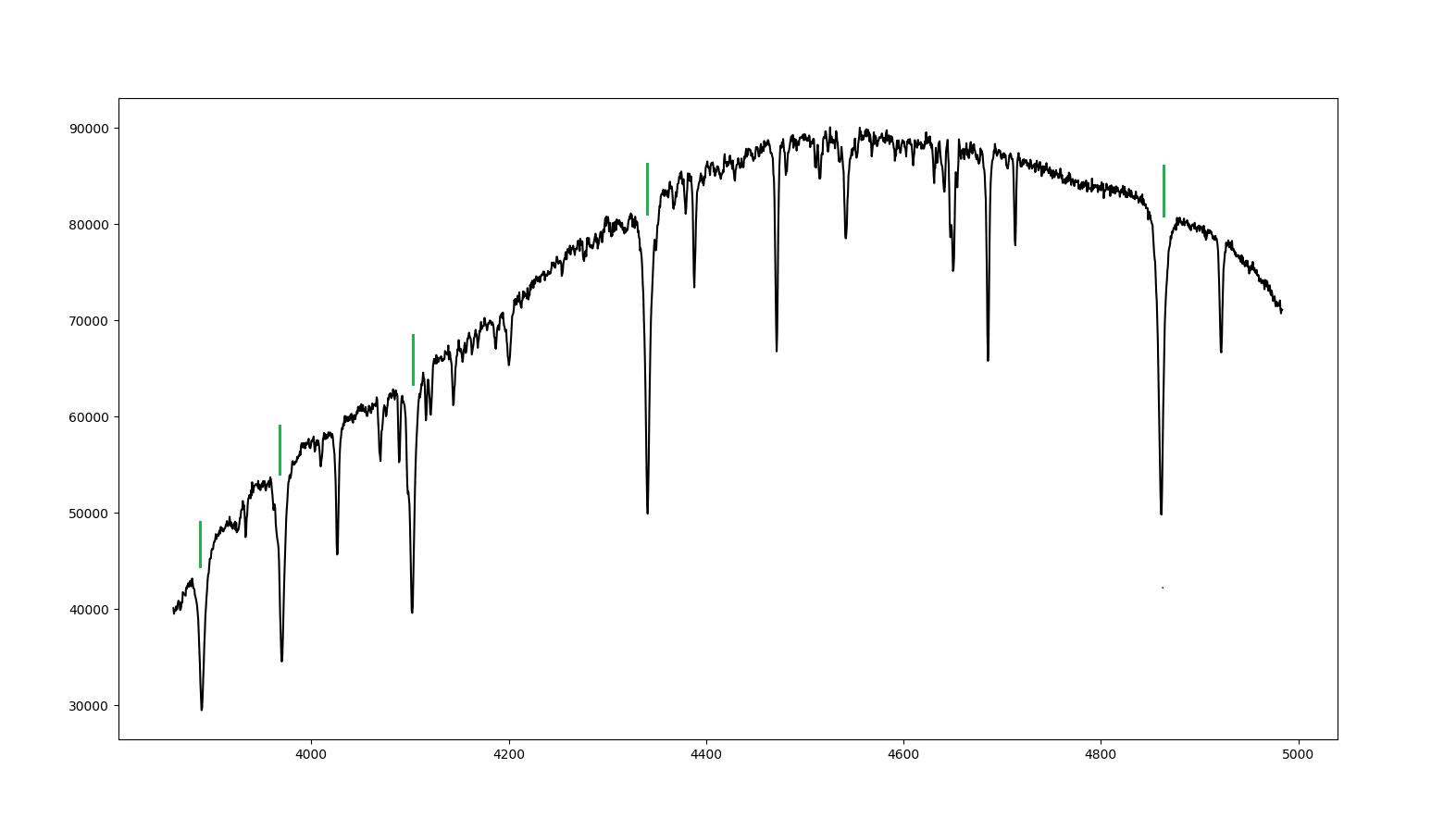
Although not a spectral classifying algorithm, *The Cannon,* a program written by a group of astronomers at MIT, presents another perspective on the problem (Ness et al. 2015). *The Cannon* was developed to find a way to efficiently determine spectral labels of large sets of data (Ness et al. 2015). It is important to mention that the term “spectral labels” refers to any label applied to the spectra at all, not specifically spectral classification (Ness et al. 2015). *The Cannon* uses a data-driven approach to determine spectral labels (Ness et al. 2015). This method involves training a classifier on a reference set of spectra, and then using the trained classifier to determine the spectral labels of unknown spectra (Ness et al. 2015). The classifier learns the characteristics of the stars’ spectra with certain labels and looks for those characteristics in the spectra of yet unlabeled stars (Ness et al. 2015). This is a very computationally efficient method of classifying spectra because it does not involve storage of large lists of metallic line strengths or predetermined line characteristics (Ness et al. 2015). *The Cannon* is very effective at determining the spectral labels of a large set of spectra with a high degree of accuracy (Ness et al. 2015). The main weakness of this approach is that it requires a sufficiently large data set to train on for it to provide accurate classifications (Ness et al. 2015). Although the program is not immediately applicable to determining classifications, there is a strong potential for adapting *The Cannon* method for use as an automated spectral classification algorithm.

In the past several years, a subfield of artificial intelligence has developed in the scientific community known as machine learning (Raschka 2015). Machine learning is the process in which an algorithm organizes data by identifying underlying patterns (Raschka 2015). There are three types of machine learning currently used for data-driven classification (Raschka 2015). We are using what is known as supervised learning, meaning we train our classification model on information that has already been classified and then use that model to classify unlabeled information (Raschka 2015). Using supervised learning means we can easily check how accurate our trained classifier is by testing it against subset of the data known as the test set (Raschka 2015). Both the training set and the test set are subsets of the larger superset of data and are comprised entirely of what is known as “strong candidates”. Strong candidates are those data that we are confident in their labelling. In this case, our strong candidates are the spectral standards. Machine learning gives us the flexibility to add more classification labels and more criteria with which to use for classification with ease. In addition, many machine learning classification algorithms can perform analysis on themselves and determine the optimal configuration of algorithm hyperparameters to achieve the most accurate and efficient version of a classifier.

So far, we have introduced stellar classification, the MK classification system, attempts at creating spectral classification algorithms and the basics of machine learning. In the following section we will discuss more specifically how spectral classification is done, how we modeled the hydrogen lines, the random forest classifier from sklearn, cross validation and hyperparameter tuning for random forests. Following that discussion, we will present the results of our classification and describe the different trials that were performed. Finally, we will discuss what our results mean, and the next steps for this project, including an introduction to the LAMOST data set and a brief outline of the characteristics we hope to add to our characterization of spectra to get a more accurate classification.

1. RESEARCH METHODS

When classifying spectra, some features are more indicative than others. The hydrogen Balmer lines (Marked with green in Figure 2) are the most indicative features of early-type stars (Gray and Corbally 2009). Shown below in Figure 3 are the Hγ absorption lines for an O9 V star, a B3 V star and an A3 V star plotted over each other. Analyzing the hydrogen lines closely, a clear progression in the strength of the lines can be seen from O9 V as the weakest and A3 V lines as the strongest. Comparing the strength of the hydrogen lines is one of the first steps used to determine spectral type by eye and thus was the first step we took in classifying spectra using a machine learning algorithm.



*Figure 2: The spectrum for HD 214680, an O9-type star with luminosity class V. The hydrogen lines are clearly marked and range from H8 to Hβ.*

Hβ

Hγ

Hδ

Hε

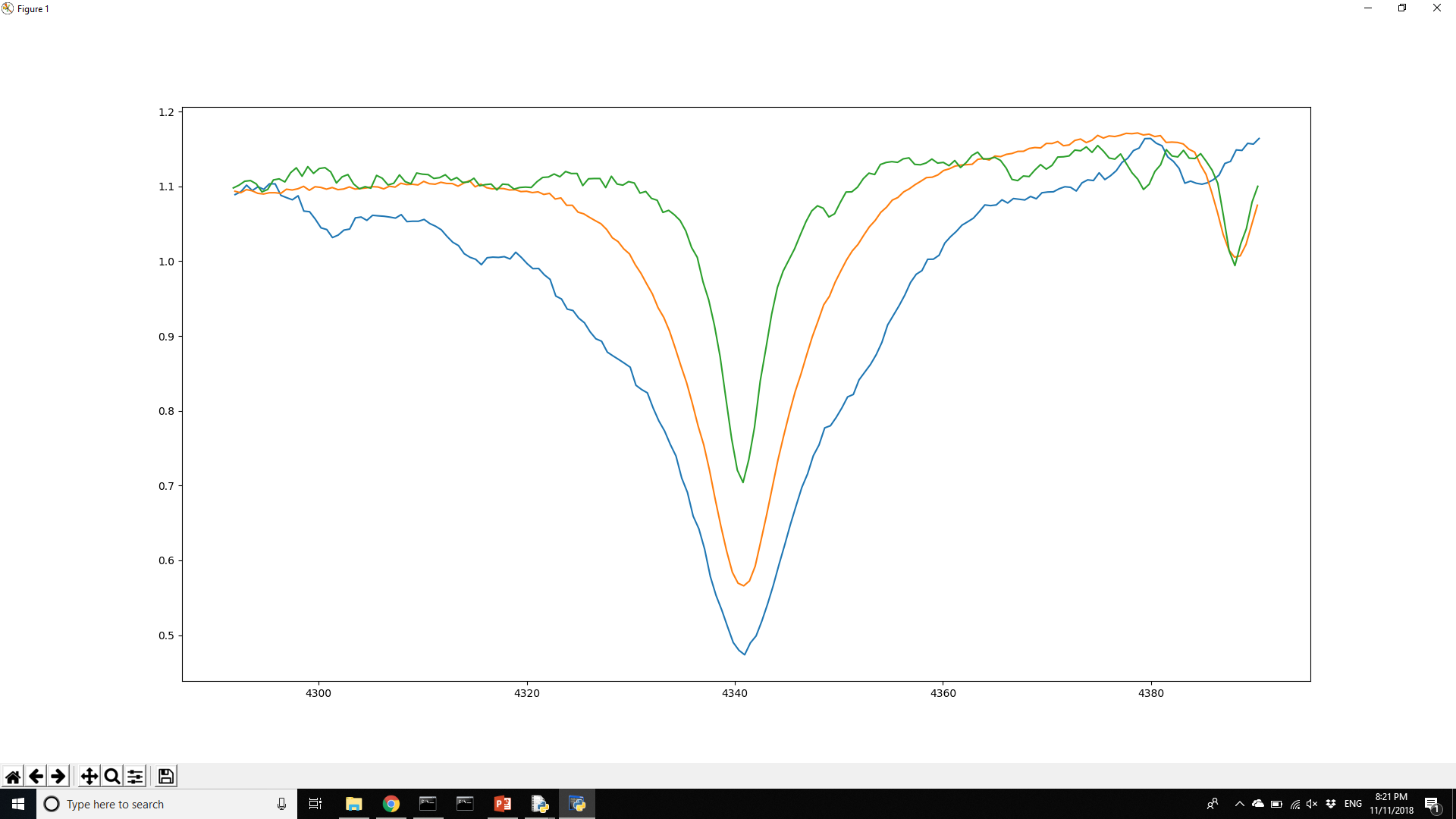
H8

Wavelength (Å)

Intensity (ADU)

HD 214680

*Figure 3: Hγ for an O9 V star (HD214680), a B3 V star (HD120315) and an A3 V star (HD023643) over-plotted. The A3 V is the strongest, B3 V is intermediary, and O9 V is the weakest as determined by the area of the hydrogen lines.*



Wavelength (Å)

Intensity (ADU)

Hγ

A3 V

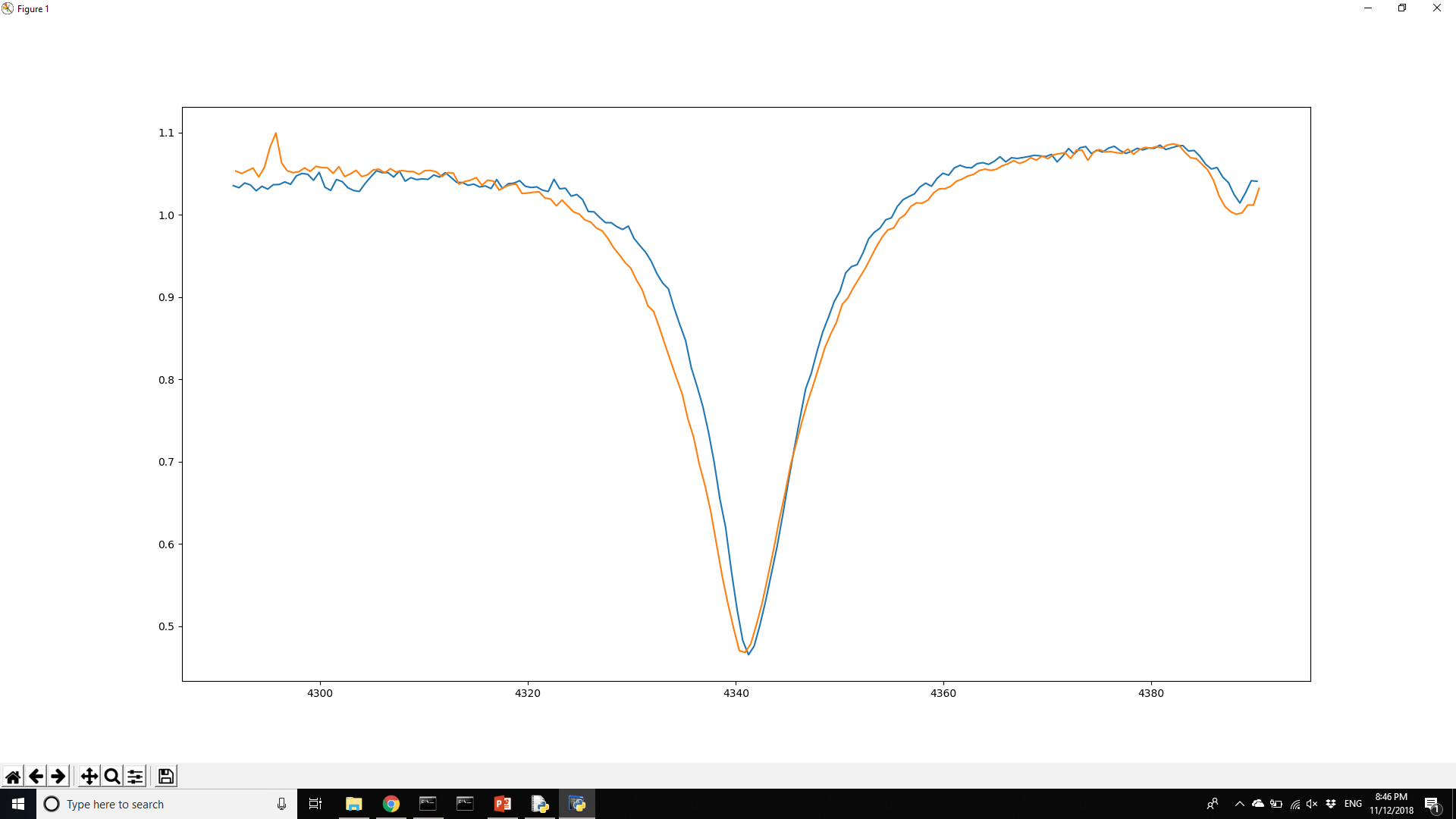
B3 V

O9 V

While the strength of the hydrogen lines is a strong indicator of the type and the width and shape of the hydrogen lines is a strong indicator of the luminosity class of the star (Gray and Corbally 2009). Figure 4 shows a close-up view of the Hγ line in the spectra of a B7 V star and a B7 III plotted over one another. The bend in the hydrogen line gets noticeably sharper from luminosity class III to V and the width of the line gets wider from III to V.

B7 III

B7 V



Hγ

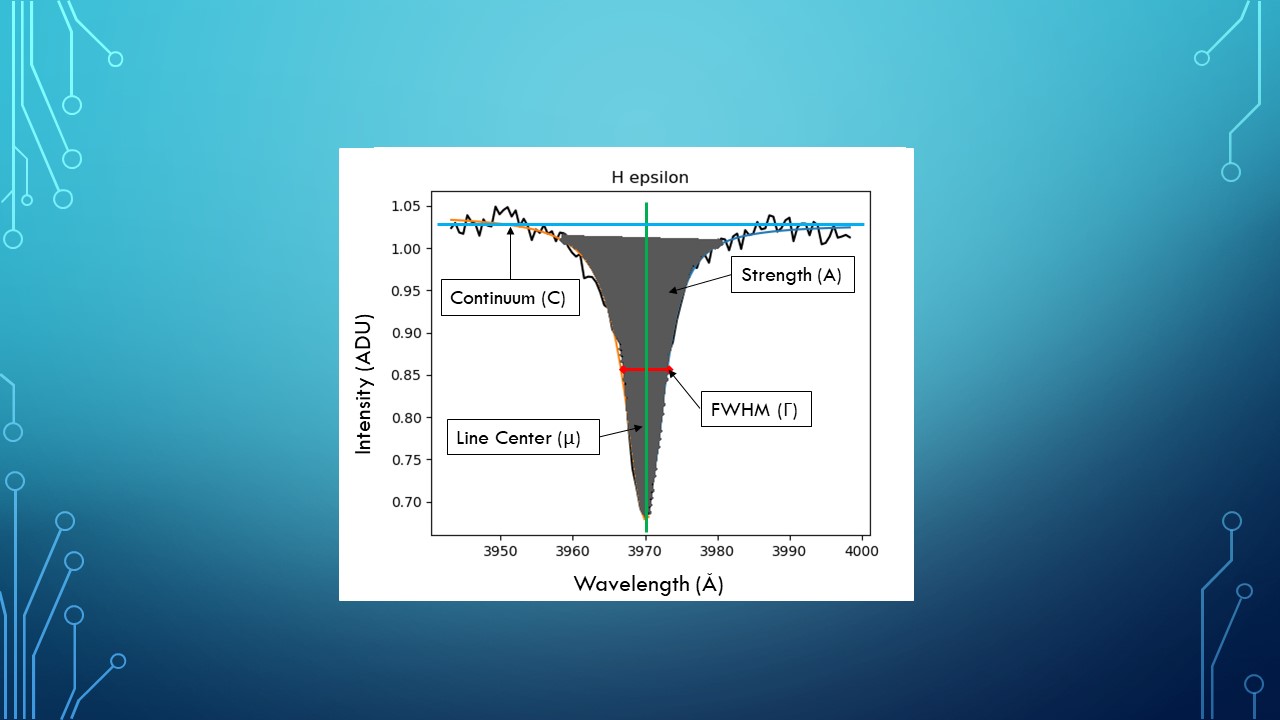
Wavelength (Å)

Intensity (ADU)

*Figure 4: Hγ for a B7 V star (HD021071) and a B7 III star (HD035497). The B7 III star is in orange and the B7 V star is in blue.*

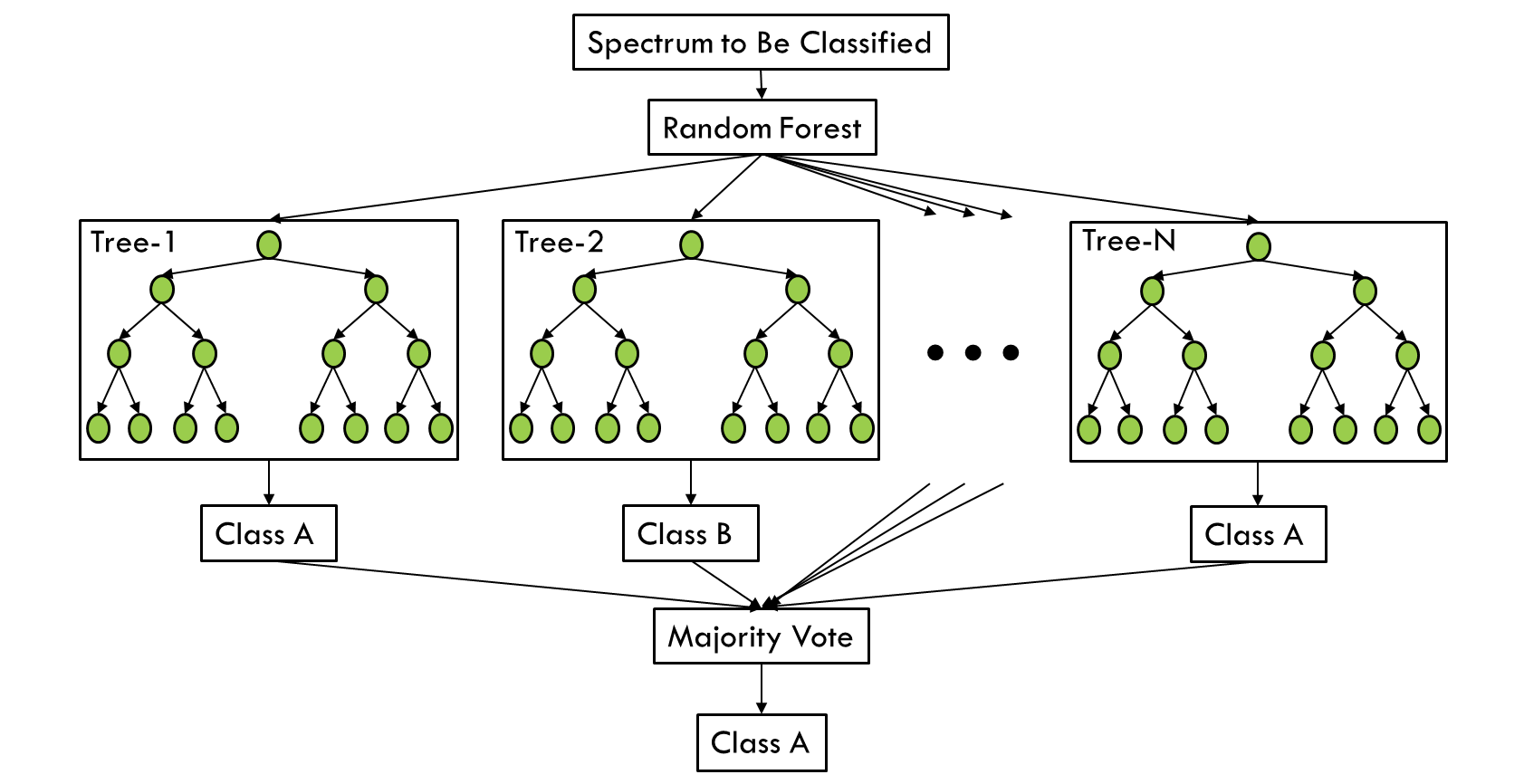
These two characteristics, the strength and the shape of the hydrogen lines are the two most important characteristics to describing the hydrogen lines (Gray and Corbally 2009). Thus, we started by fully describing these characteristics of the hydrogen lines. To transform the “by-eye” analysis usually done by astronomers into numeric values for the algorithm to use, we had to find a way to describe the hydrogen lines quantitatively. Conveniently, the strength and shape of the hydrogen Balmer lines can be described by a single equation known as the Lorentzian probability distribution (a.k.a. Cauchy distribution) (Bevington and Robinson 2010). The Lorentzian probability distribution is described using the equation below:

+ C

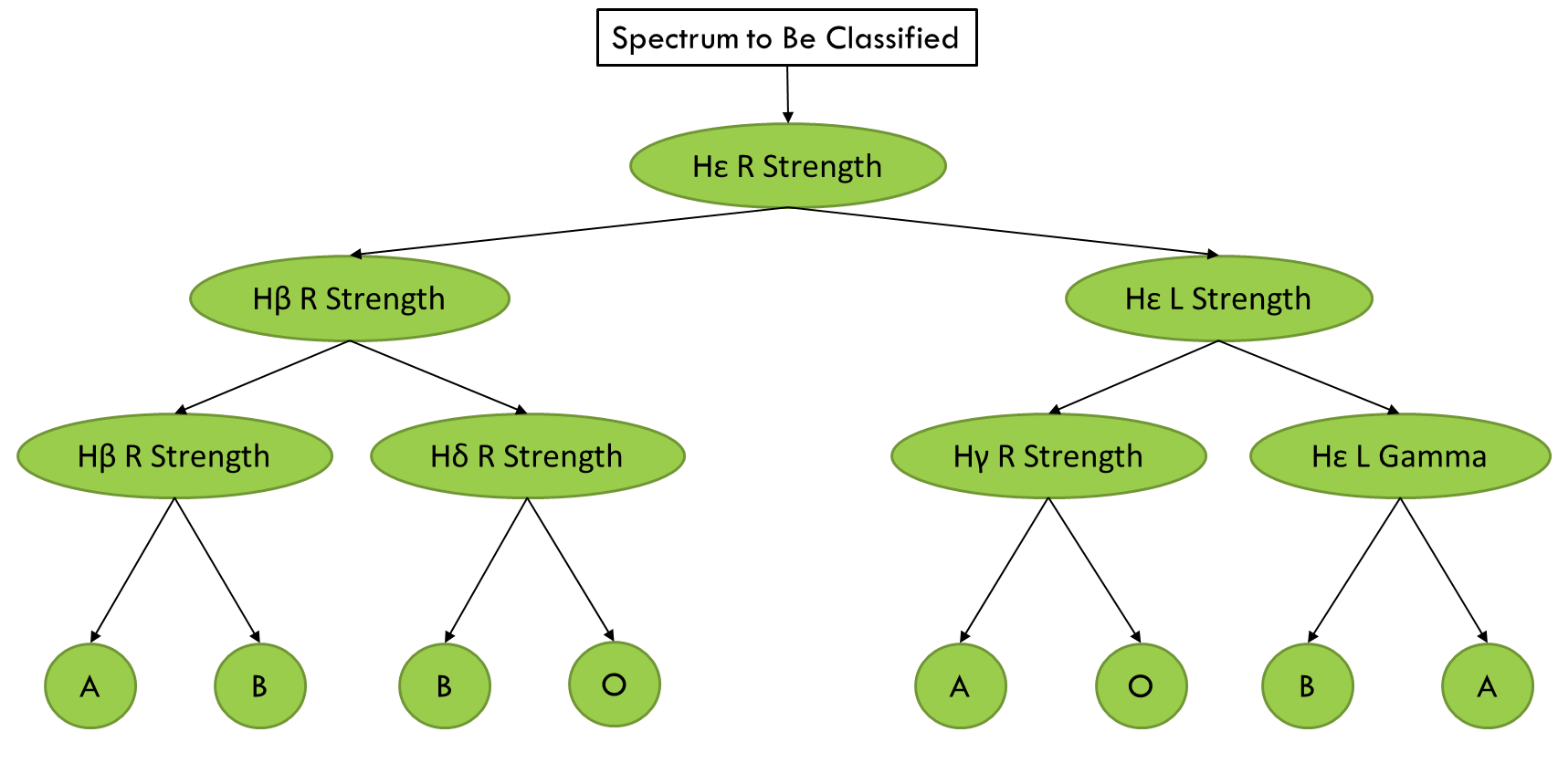
 Where Γ is the half-width of the line, μ is the center of the line, A is the strength of the line and C is the height of the continuum (See Figure 5) (Bevington and Robinson 2010). Using this equation and the curve\_fit function from the SciPy package, we can fully describe the hydrogen lines quantitatively for each spectrum (Jones, et al. 2001).

*Figure 5: An H Epsilon hydrogen line with the variables in defined in the Lorenztian probability distribution shown on the graph.*

Once we had the hydrogen lines described quantitatively, we chose our unsupervised learning algorithm. We used a random forest classifier from SciKit-Learn as our machine learning algorithm because of the ease of testing the accuracy algorithm (Pedregosa et al. 2011). A random forest classifier is comprised of a number of randomly generated decision tree classifiers (Pedregosa et al. 2011). A decision tree classifier (Figure 6) classifies data by asking a series of questions about the data to determine the patterns in the data. The tree decides which questions to ask based on a characteristic of each feature in the data table called information gain (Raschka 2015). Information gain is a measure of how decisive a certain feature is (Raschka 2015). For example, a feature that results in a 50/50 split in the data is not very decisive and does not have a very high information gain, but a feature that causes a 90/10 split in the data is very decisive and has a high information gain (Raschka 2015). So, a decision tree grows itself based on a set of decisive questions, meant to categorize data into separate groups as quickly and accurately as possible (Raschka 2015). A decision tree is highly dependent on the data that it is given to train on. Thus, a group of randomly generated decision trees that are trained on random subsections of the training data and are all different (Raschka 2015). The random forest classifier (Figure 7) allows its decision trees to vote on the classification of the piece of data and decides the classification by a majority vote (Raschka 2015).



*Figure 7: Diagram of a random forest classifier. Each tree votes on the spectrum and the majority vote is what the forest determines as a classification.*



*Figure 6: Diagram of decision tree. Each oval is a decision node, each circle is a classification node. The higher a node is in the tree the larger its information gain.*

To analyze the quality of our model we used k-fold cross validation (Pedregosa et al. 2011). K-fold cross validation (KCV) (Figure 8) is the process of splitting the strong candidates into k different sections and then using one section for testing and the rest for training and then training the model k times (Raschka 2015). We set k equal to 10 for our KCV, meaning the list of strong candidates is split into 10 sections. KCV gives a more accurate depiction of the accuracy of the model by providing an average accuracy for the model over the entire list of strong candidates as opposed to the accuracy of the model for a subsection of the strong candidates (Raschka 2015). Once we have an accuracy rating for the model itself, we can do what is call hyperparameter tuning. Hyper parameter tuning is the process of changing the settings of the classifier to find the optimal settings and then retraining the “tuned” model on the entire training set (Raschka 2015).

*Figure 8: Diagram detailing K-fold cross validation. The process shown here has a K value of 5 whereas our process has a K value of 10.*

1. RESULTS

Our data set was taken at the Adams Observatory over the past 3 years by David Whelan. As mentioned before, the set of spectra we are using consists only of O-, B-, and A-type stars and, at the time this paper is being written, consists of 1025 spectra. Of those 1025 we have identified 96 strong candidates to be used in classification. The strong candidates are those spectra that are either spectral standards or spectra that we are extremely confident in their classification. The list of strong candidates and their “by eye” classifications can be found in Appendix A of this document. The machine learning classification was done with 4 different versions of the labels. The classifier was trained on labels that contained only the letter type of the star (i.e. A, B, or O), labels that contained the letter type and whole number subtype (i.e. A1, A3, B8), labels that contained the full spectral type of the star, and labels that contained the entire classification of the star. These different labelling schemes were used to observe any effects generalizing the data might have. The different classification schemes can be found in Appendix B. The sampling size is the ratio of training set to test set. The accuracy rating is the percentage of spectra in the training set classified correctly by the trained model. An accuracy rating of 70% indicates that 70% of the spectra in the test set were classified correctly using that model.

The following table shows the information recorded about each training of the classifier that was performed.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Date | Classifier | Accuracy | Type Version | Sampling Size | Comments |
| 10/18/2018 | Random Forest Classifier | 95% | Alpha Only | 80/20 | First Success!!! |
| 10/18/2018 | Random Forest Classifier | 70% | Whole Num | 80/20 | Less spectra of each category results in a more inaccurate fit |
| 10/18/2018 | Random Forest Classifier | 60% | No Lumin | 80/20 | Again, less spectra results in more inaccurate, although not as big of a difference as I was expecting here |
| 10/18/2018 | Random Forest Classifier | 60% | Complete | 80/20 | I have to believe there is some overfitting here. There can't not be. There are 51 unique classifications being applied to not even double that. |
| 10/23/2018 | RFC Avg using K-Fold Cross-validation | 79% | Alpha Only | 90/10 | This feels like a more accurate representation, I'm interested to see how the rest of the fits go |
| 10/23/2018 | RFC Avg using K-Fold Cross-validation | 28% | Whole Num | 90/10 | This makes much more sense. |
| 10/23/2018 | RFC Avg using K-Fold Cross-validation | 23% | No Lumin | 90/10 |  |
| 10/23/2018 | RFC Avg using K-Fold Cross-validation | 24% | Complete | 90/10 |  |

1. DISCUSSION

Using the more accurate statistic from the KCV set, our classifier is currently predicting with 79% accuracy on the generalized set of labels and with 23-28% accuracy on the more specific sets of labels. This indicates that we need more data. The generalized labels were used to simulate what kind of statistics we would achieve if we had triple the amount of data and 79% accuracy is quite good considering we are fitting solely based on the Lorentz fits of the hydrogen lines. The much lower accuracy of the Whole Number, No Luminosity Class and Complete Label trials indicate that we are not providing the classifying algorithm with enough data to make appropriate classifications. It is impossible for the algorithm to determine a pattern if it only has one or two examples of a certain label.

There are three ways to solve this problem that we are currently pursuing: borrowing data from another astronomy group, manufacturing more data and taking more spectra at the Adams Observatory to add to our training set.

We are looking to use the data set taken by the Large Sky Multi-Object Fiber Spectroscopic Telescope (LAMOST) used by the Chinese Academy of Sciences (“LAMOST DR1” 2015). The LAMOST survey has taken spectra from millions of stars and galaxies and would have more than enough data for use to successfully create a well-trained model (“LAMOST DR1” 2015). Additionally, the LAMOST data set has more types of spectra other than just the O-, B- and A-types that are collected at the Adams Observatory (“LAMOST DR1”). The only possible downside of using this data is that it may be in an incompatible format to use with our algorithm. Our algorithm requires the data to be wavelength calibrated before being input into the program. Wavelength calibration is already a nontrivial process with the data we have at Austin College, having to calibrate the spectra for millions of spectra would take a prohibitively long time. However, it is possible we could use an automated wavelength calibration algorithm that was developed over the summer and that could alleviate that issue. We are currently in the process of asking for permission to use this data set.

Another possible solution we are pursuing is manufacturing new data based off old data, a practice that is quite common in machine learning when faced with a lack of data. We would add gaussian noise to the spectra we already have to make new spectra that are still strongly classified and can be used to train the model. Using this approach means our new data would be a more realistic representation of what the classifier should expect from the actual data as opposed to creating new models from scratch. This approach would add more strong candidates to our list that we would not have to do any additional preprocessing for and would have an instrument profile consistent with the instrument profile of the spectra taken at the Adams Observatory.

The final solution we are pursuing is the collection of more data at the Adams Observatory. This solution has the same benefits of the manufacturing of data with the additional benefit that we know that the classifier is not making decisions off potentially unrealistic representations of spectra. There are no real downsides to this approach, but the collection of data is quite slow compared to the other methods and highly dependent on the weather here in Sherman, Texas. Additionally, some of the spectra that we would like to take are of stars that are not currently in the night sky and will not be again until the summer after this project concludes.

Once we have obtained more data, the next step for the analysis of the spectra is adding in more metallic lines. The next lines we are going to analyze are the helium I and helium II absorption lines, as they are also quite characteristic of early type stars (Gray and Corbally 2009). The shape of metallic line wings is not as important as the strength of the metallic lines, and so we will use a gaussian model to describe the strength of the metallic lines. Once we have classified using the helium lines in addition to the hydrogen lines, depending on performance we may use the silicon lines as well. If we get access to the LAMOST data set we will be able to include characteristics that are useful to classify late-type stars as well as early-type stars but for now we are focusing on early type stars.

1. REFERENCES

Bevington, Philip R., and D. Keith. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, 2010.

Gray, R. O., and C. J. Corbally. “An Expert Computer Program For Classifying Stars On The Mk Spectral Classification System.” *The Astronomical Journal*, vol. 147, no. 4, 2014, p. 80., doi:10.1088/0004-6256/147/4/80.

Gray, Richard O., and Christopher J. Corbally. *Stellar Spectral Classification*. Princeton University Press, 2009.

Gray, R. O., et al. “Lamost Observations In Thekeplerfield: Spectral Classification With The Mkclass Code.” *The Astronomical Journal*, vol. 151, no. 1, 2015, p. 13., doi:10.3847/0004-6256/151/1/13.

“Hertzsprung-Russell Diagram.” *COSMOS - The SAO Encyclopedia of Astronomy | COSMOS*, Swinburne University, 2015, astronomy.swin.edu.au/cosmos/H/Hertzsprung-Russell Diagram.

Jones, Eric, et al. “SciPy: Open Source Scientific Tools for Python.” *SciPy: Open Source Scientific Tools for Python*, 2001, [www.scipy.org/](http://www.scipy.org/).

“LAMOST DR1” *LAMOST DR1 - Home*, Chinese Academy of Sciences, 2015, dr1.lamost.org/.

Morgan, W. W., and P. C. Keenan. “Spectral Classification.” *Annual Review of Astronomy and Astrophysics*, vol. 11, no. 1, 1973, pp. 29–50., doi:10.1146/annurev.aa.11.090173.000333.

Pedregosa, et al. “Scikit-Learn: Machine Learning in Python.” *JLMR*, vol. 12, 2011, pp. 2825–2830.

Raschka, Sebastian. *Python Machine Learning*. Packt Publishing Limited, 2015.

APPENDIX A

The following is a list of the 96 strong candidates used our classification algorithm.

|  |  |  |  |
| --- | --- | --- | --- |
| Spectral Type | Luminosity Class | Additional Label | Name and Date of Star |
| O9 | III | - | HD024431\_20171211.txt |
| O9 | III | n | HD028446A\_20171211.txt |
| O9 | V | - | HD214680\_20171009.txt |
| O9.2 | IV | nn | HD149757\_20180618.txt |
| O9.5 | III | nn | HD093521\_20180303.txt |
| O9.5 | IV | - | HD202214\_20170922.txt |
| O9.5 | IV-V | - | HD206183\_20171025.txt |
| O9.5 | V | - | HD034078\_20171211.txt |
| O9.7 | III | - | HD189957\_20170910.txt |
| O9.7 | IV | - | HD209339\_20171009.txt |
| B0 | V | - | HD036512\_20171211.txt |
| B0.2 | III | - | HD006675\_20171126.txt |
| B0.2 | IV | - | HD036822\_20171211.txt |
| B0.2 | V | - | HD149438\_20160607.txt |
| B0.2 | V | - | HD149438\_20170524.txt |
| B0.5 | III | - | HD218376\_20171013.txt |
| B0.5 | IV | - | HD034816\_20171211.txt |
| B0.5 | V | - | HD036960\_20171211.txt |
| B0.7 | III | - | HD024760\_20170211.txt |
| B1 | III | - | HD023180\_20171126.txt |
| B1 | V | - | HD144470\_20160607.txt |
| B1 | V | - | HD144470\_20170524.txt |
| B1.5 | III | - | HD214993\_20171013.txt |
| B1.5 | V | - | HD154445\_20160607.txt |
| B1.5 | V | - | HD154445\_20170524.txt |
| B2 | III | - | HD030836\_20171211.txt |
| B2 | III | - | HD035468\_20171211.txt |
| B2 | III | - | HD035468\_20180105.txt |
| B2 | IV | - | HD000886\_20171009.txt |
| B2 | IV | - | HD016582\_20171126.txt |
| B2 | V | - | HD145482\_20180618.txt |
| B2.5 | III | - | HD207330\_20171009.txt |
| B2.5 | V | - | HD148605\_20160607.txt |
| B2.5 | V | - | HD148605\_20170524.txt |
| B2.5 | V | - | HD214432\_20171009.txt |
| B3 | III | - | HD021483\_20170109.txt |
| B3 | IV | - | HD037711\_20180303.txt |
| B3 | V | - | HD020365\_20171126.txt |
| B3 | V | - | HD032630\_20170109.txt |
| B3 | V | - | HD074280\_20180302.txt |
| B3 | V | - | HD120315\_20150624.txt |
| B3 | V | - | HD120315\_20160607.txt |
| B3 | V | - | HD120315\_20170524.txt |
| B3 | V | - | HD120315\_20170608.txt |
| B3 | V | - | HD120315\_20180607.txt |
| B3 | V | - | HD120315\_20180613.txt |
| B3 | V | - | HD120315\_20180625.txt |
| B5 | III | - | HD034503\_20171211.txt |
| B5 | IV | - | HD147394\_20150625.txt |
| B5 | IV | - | HD147394\_20160607.txt |
| B5 | IV | - | HD147394\_20170524.txt |
| B5 | V | - | HD034759\_20171211.txt |
| B6 | IV | - | HD023338\_20171126.txt |
| B7 | III | - | HD035497\_20171211.txt |
| B7 | IV | n | HD023288\_20171126.txt |
| B7 | V | - | HD021071\_20170109.txt |
| B8 | III | - | HD023850\_20171126.txt |
| B8 | V | - | HD019356\_20170104.txt |
| B8 | V | - | HD019356\_20170109.txt |
| B8 | V | n | HD023324\_20171126.txt |
| B9 | III | - | HD176437\_20170922.txt |
| B9 | IV | - | HD196867\_20170910.txt |
| B9 | V | a | HD016046\_20171126.txt |
| B9.5 | V | a | HD023873\_20171211.txt |
| A0 | III | n | HD061931\_20180302.txt |
| A0 | III | - | HD087887\_20170524.txt |
| A0 | III | - | HD123299\_20170524.txt |
| A0 | IV | - | HD210419\_20171009.txt |
| A0 | V | n | HD023629\_20171211.txt |
| A0 | V | a | HD071155\_20180302.txt |
| A0 | V | an | HD103287\_20170524.txt |
| A0 | V | a | HD172167\_20150609.txt |
| A0 | V | a | HD172167\_20150915.txt |
| A0 | V | a | HD172167\_20150917.txt |
| A0 | V | a | HD172167\_20150929.txt |
| A0 | V | a | HD172167\_20170524.txt |
| A0 | V | a | HD172167\_20170910.txt |
| A0 | V | a | HD172167\_20170915.txt |
| A0 | V | a | HD172167\_20170922.txt |
| A0 | V | a | HD172167\_20171012.txt |
| A0 | V | a | HD172167\_20180702.txt |
| A0 | V | a | HD172167\_20180703.txt |
| A0 | V | a | HD201184\_20170922.txt |
| A1 | IV | nn | HD073262\_20180302.txt |
| A1 | IV | - | HD139006\_20150623.txt |
| A1 | IV | - | HD139006\_20180607.txt |
| A1 | IV | - | HD139006\_20180618.txt |
| A1 | IV | - | HD216735\_20171013.txt |
| A1 | V | a | HD009132\_20171126.txt |
| A2 | III | - | HD056169\_20180410.txt |
| A2 | IV | - | HD097277\_20170524.txt |
| A2 | IV | - | HD141003\_20170524.txt |
| A3 | IV | n | HD033111\_20171211.txt |
| A3 | V | - | HD023643\_20171211.txt |
| A3 | V | a | HD102647\_20170524.txt |
| A3 | V | a | HD216956\_20171025.txt |

APPENDIX B

The following is a tabling showing the different labelling schemes.

|  |  |  |  |
| --- | --- | --- | --- |
| SpTypeComplete | SpTypeNoLumin | SpTypeWholeNumber | SpTypeAlphaOnly |
| O9III | O9 | O9 | O |
| O9III | O9 | O9 | O |
| O9V | O9 | O9 | O |
| O9.2IV | O9.2 | O9 | O |
| O9.5III | O9.5 | O9 | O |
| O9.5IV | O9.5 | O9 | O |
| O9.5IV-V | O9.5 | O9 | O |
| O9.5V | O9.5 | O9 | O |
| O9.7III | O9.7 | O9 | O |
| O9.7IV | O9.7 | O9 | O |
| B0V | B0 | B0 | B |
| B0.2III | B0.2 | B0 | B |
| B0.2IV | B0.2 | B0 | B |
| B0.2V | B0.2 | B0 | B |
| B0.2V | B0.2 | B0 | B |
| B0.5III | B0.5 | B0 | B |
| B0.5IV | B0.5 | B0 | B |
| B0.5V | B0.5 | B0 | B |
| B0.7III | B0.7 | B0 | B |
| B1III | B1 | B1 | B |
| B1V | B1 | B1 | B |
| B1V | B1 | B1 | B |
| B1.5III | B1.5 | B1 | B |
| B1.5V | B1.5 | B1 | B |
| B1.5V | B1.5 | B1 | B |
| B2III | B2 | B2 | B |
| B2III | B2 | B2 | B |
| B2III | B2 | B2 | B |
| B2IV | B2 | B2 | B |
| B2IV | B2 | B2 | B |
| B2V | B2 | B2 | B |
| B2.5III | B2.5 | B2 | B |
| B2.5V | B2.5 | B2 | B |
| B2.5V | B2.5 | B2 | B |
| B2.5V | B2.5 | B2 | B |
| B3III | B3 | B3 | B |
| B3IV | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B3V | B3 | B3 | B |
| B5III | B5 | B5 | B |
| B5IV | B5 | B5 | B |
| B5IV | B5 | B5 | B |
| B5IV | B5 | B5 | B |
| B5V | B5 | B5 | B |
| B6IV | B6 | B6 | B |
| B7III | B7 | B7 | B |
| B7IV | B7 | B7 | B |
| B7V | B7 | B7 | B |
| B8III | B8 | B8 | B |
| B8V | B8 | B8 | B |
| B8V | B8 | B8 | B |
| B8V | B8 | B8 | B |
| B9III | B9 | B9 | B |
| B9IV | B9 | B9 | B |
| B9V | B9 | B9 | B |
| B9.5V | B9.5 | B9 | B |
| A0III | A0 | A0 | A |
| A0III | A0 | A0 | A |
| A0III | A0 | A0 | A |
| A0IV | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A0V | A0 | A0 | A |
| A1IV | A1 | A1 | A |
| A1IV | A1 | A1 | A |
| A1IV | A1 | A1 | A |
| A1IV | A1 | A1 | A |
| A1IV | A1 | A1 | A |
| A1V | A1 | A1 | A |
| A2III | A2 | A2 | A |
| A2IV | A2 | A2 | A |
| A2IV | A2 | A2 | A |
| A3IV | A3 | A3 | A |
| A3V | A3 | A3 | A |
| A3V | A3 | A3 | A |
| A3V | A3 | A3 | A |