

Classifying Stellar Spectra

By Edith Johnston

Table of Contents

1. Science Intro
2. Data and Model
3. Model Performance
4. Next Steps



Stellar Spectral Types



O



B



A



F



G

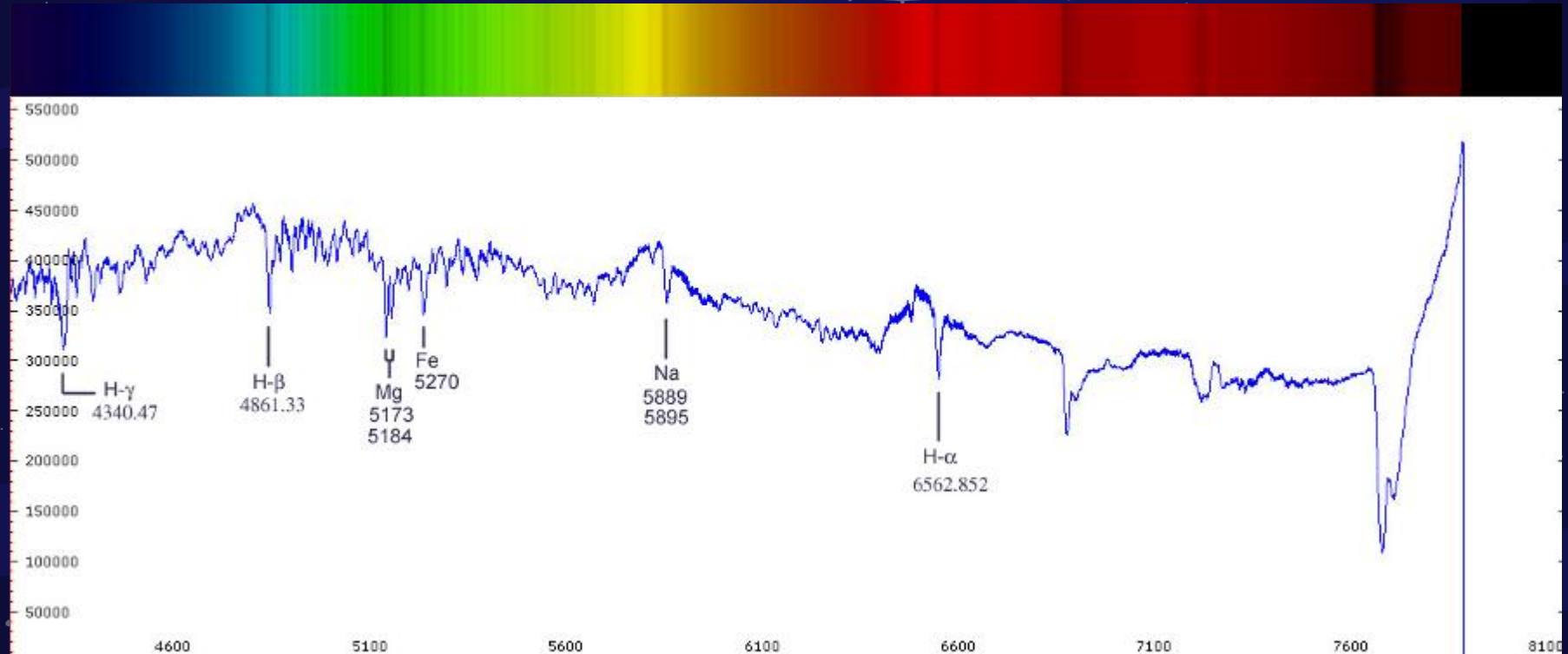


K



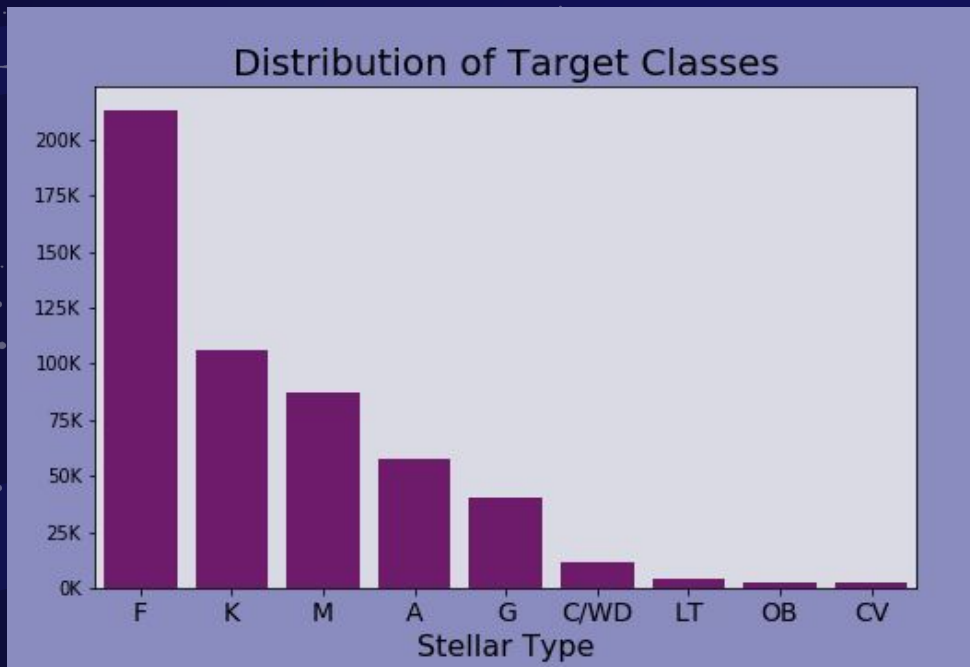
M

Stellar Spectra



Tools used: Imblearn,
Sci-kit learn, SciServer,
Pandas

Class Distribution



- ★ Largest 5 Classes:
Random Undersampling
- ★ Smallest 4 Classes:
Random Oversampling

Advantages of This Model:

Model

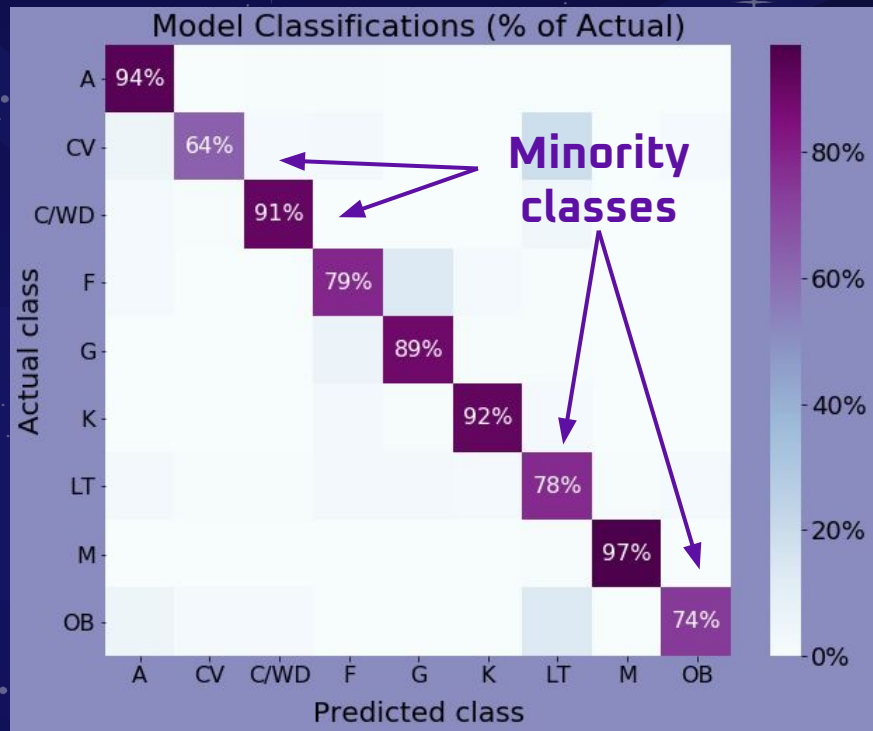
**Histogram-based
Gradient Boosting
Classification Tree**

- ★ Scales well to more features
- ★ Faster for large sample sizes
- ★ Missing value handling

Tools used: Pandas,
Sci-kit learn, SciServer,
XGBoost

Model Performance

Balanced accuracy
score:
0.84



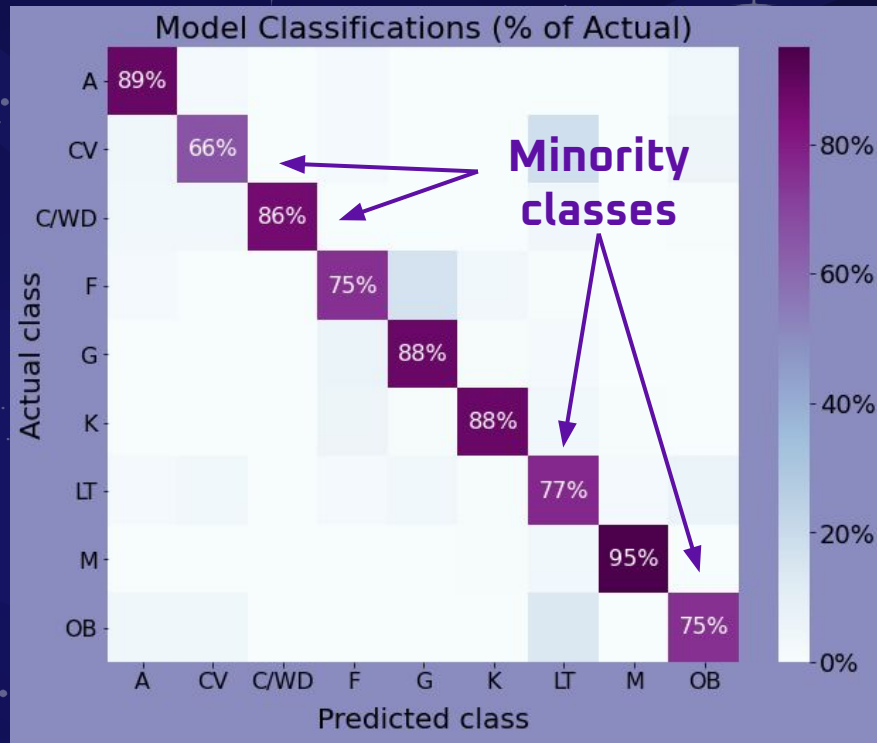
Correct classifications are on the diagonal

Hist-Gradient Boosted

Tools used: Tensorflow,
Sci-kit learn, Pandas,
Google Colab

Neural Net Model

Balanced accuracy
score:
0.83



Correct classifications are on the diagonal

Next Steps

Adding Capability

- ★ Subclasses
- ★ Non-stellar objects
 - Galaxies
 - Quasars

Improving Performance

- ★ Cross reference other sources
- ★ Train iteratively with resampled data



Thank You For Your Time



Project by: Edith Johnston

Github: <https://github.com/edithalice>

LinkedIn: www.linkedin.com/in/edith-johnston

The Fine Print

This research makes use of the SciServer science platform (www.sciserver.org).

SciServer is a collaborative research environment for large-scale data-driven science. It is being developed at, and administered by, the Institute for Data Intensive Engineering and Science at Johns Hopkins University. SciServer is funded by the National Science Foundation through the Data Infrastructure Building Blocks (DIBBs) program and others, as well as by the Alfred P. Sloan Foundation and the Gordon and Betty Moore Foundation.

Appendix

The background is a dark blue space scene. It features a central light purple rectangular area. Within this area, there is a pink planet with a ring in the upper left, a pink crescent moon in the lower right, and several white stars of varying sizes. Four purple shooting stars with long tails are also present, two on the left and two on the right, pointing towards the center.

What's the Point of Classifying?

- ★ Different behavior
- ★ Filtering searches
 - Potentially habitable planets can be found orbiting F or G stars

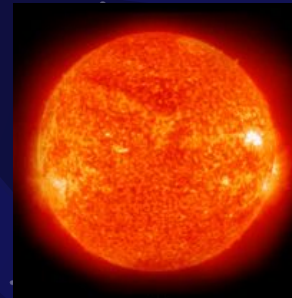


My Project Goal

Too many stars!



Automation



Data Source:

[Sloan Digital Sky
Survey](#)

Data Structure

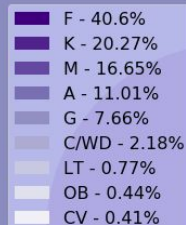
Features

- 71 features with 4 sub features each
- Each feature represent elemental line index
- Sub features: global continuum fit, local value fit, error, pixel quality mask

Target

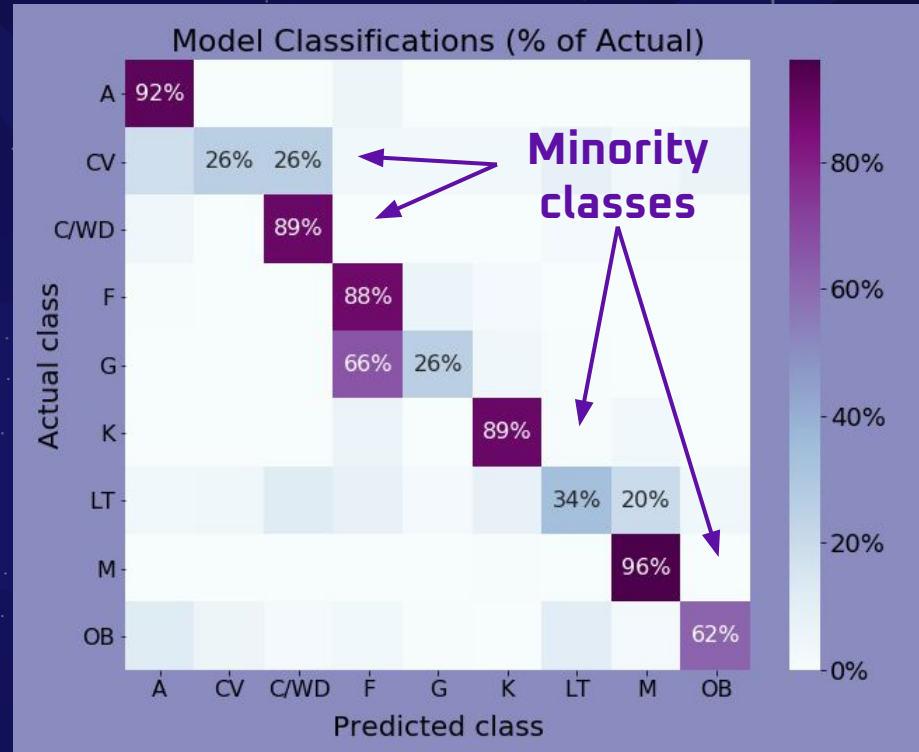
- Initially, 21 classes
- After grouping by subclass, only 9 classes
- Highly imbalanced: 5 of the 9 classes contain >96% of samples

Distribution of Target Classes

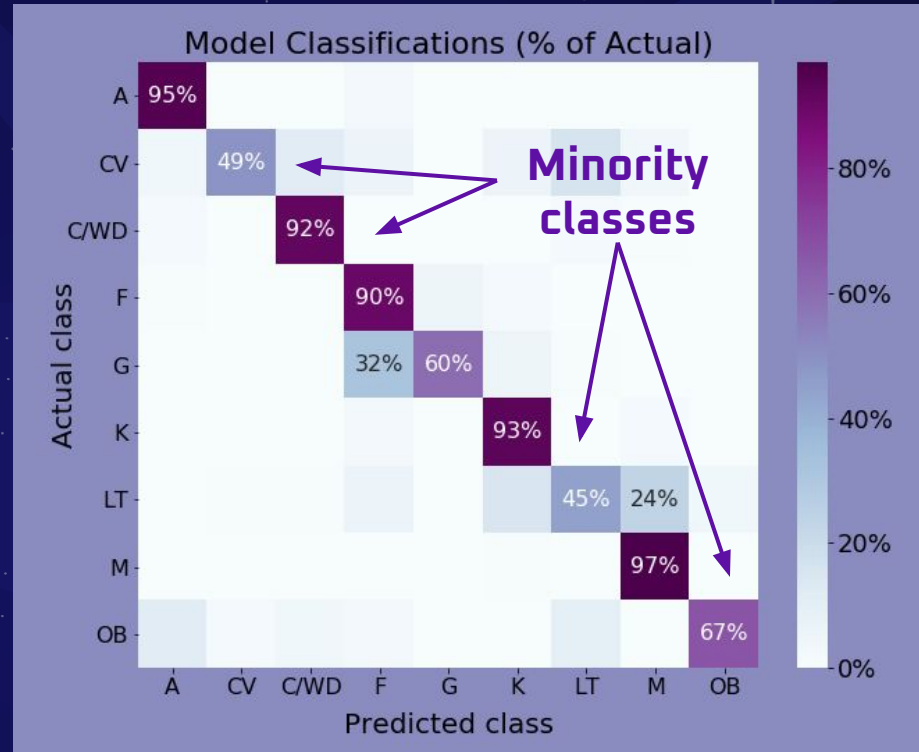


**Model
For Comparison:
Decision Tree**

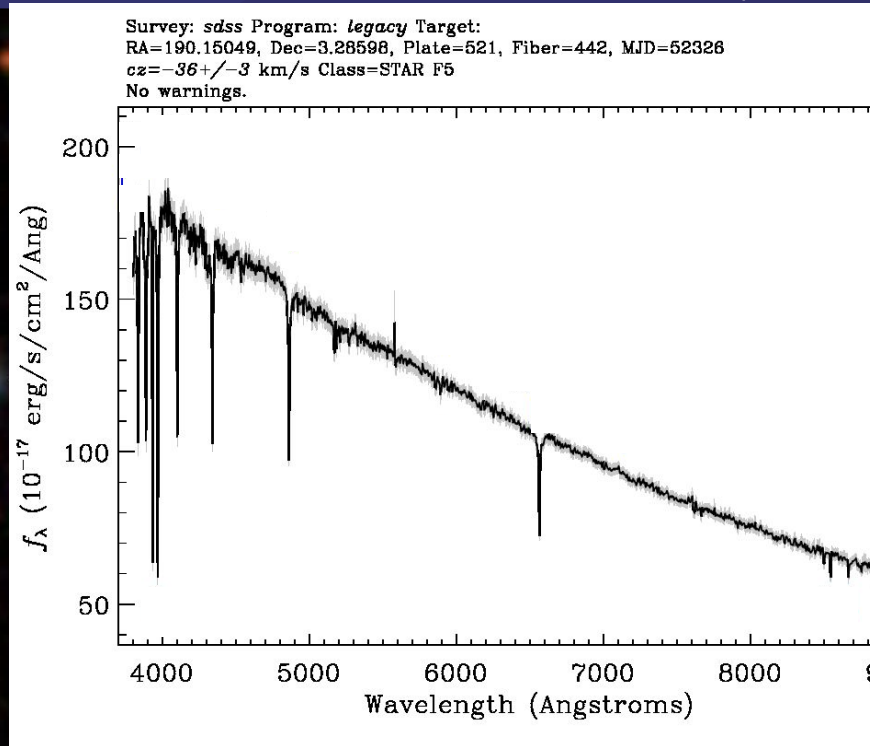
**Balanced accuracy
score:
0.67**



**Model
For Comparison:
Random Forest
Balanced accuracy
score:
0.77**



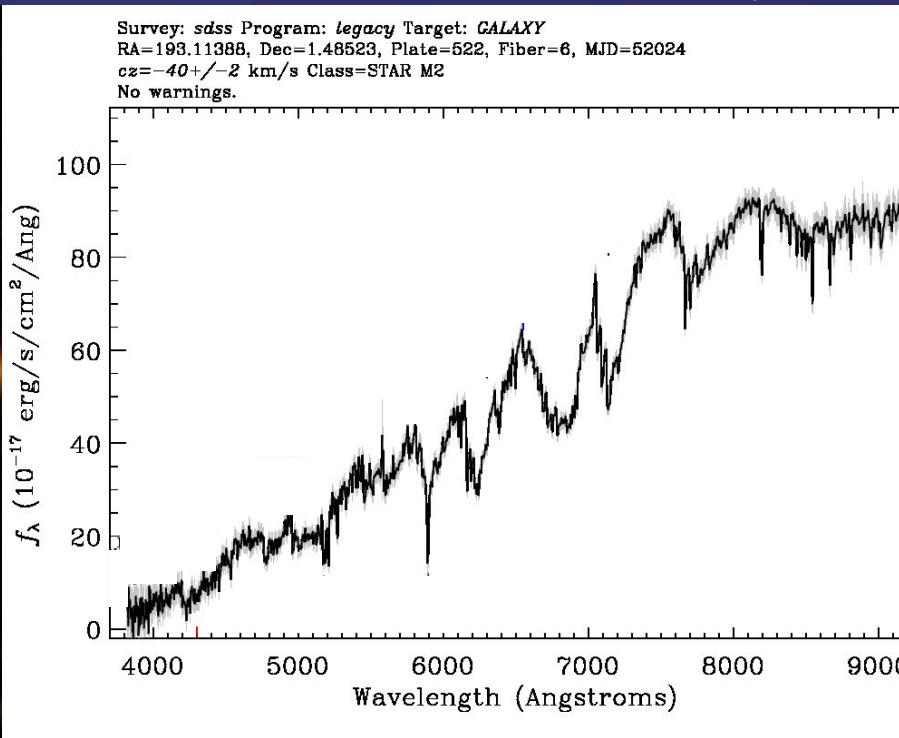
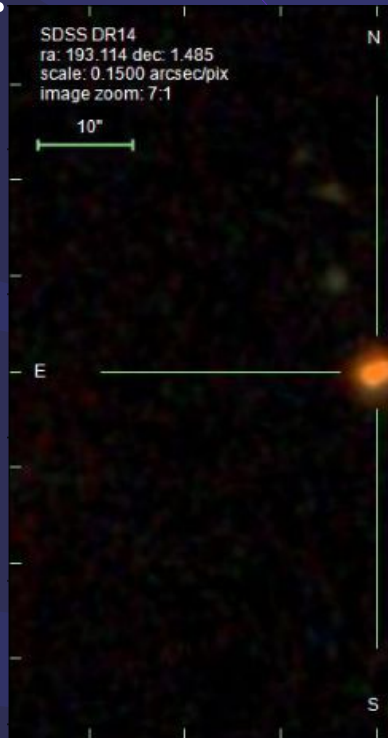
Sample Pipeline



F-type star (F)

Mass: 0.8–1.4 M_{\odot}
Radius: 1.15–1.4 R_{\odot}
Luminosity: 1.5–5 L_{\odot}
Temperature: 6,000–7,600 K

Sample Pipeline 2



Red dwarf (M)

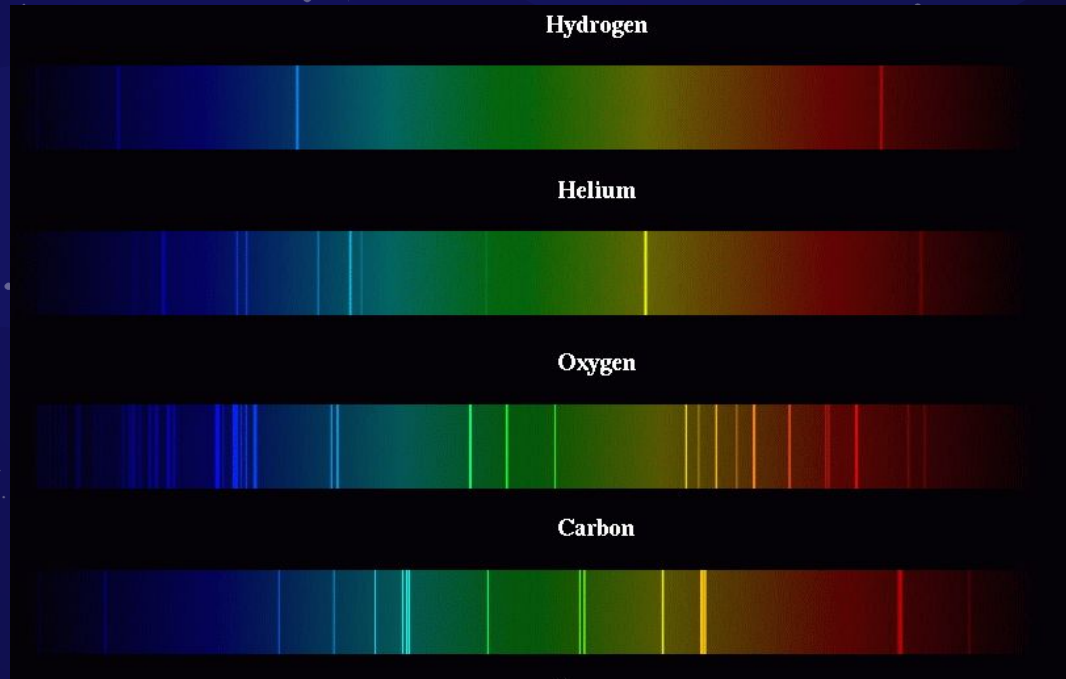
Mass: 0.075–0.5 M_{\odot}
Radius: 0.08–0.45 R_{\odot}
Luminosity: 0.015–0.08 L_{\odot}
Temperature: 2,300–4,000 K

The background is a dark blue space scene. It features numerous small white dots representing stars. Several bright, diagonal streaks of light, colored in shades of purple and blue, represent shooting stars or meteors. In the lower-left area, there is a pink planet with a thin, light blue ring. In the upper-right area, there is a pink crescent moon. The central text is contained within a dark blue rectangular box.

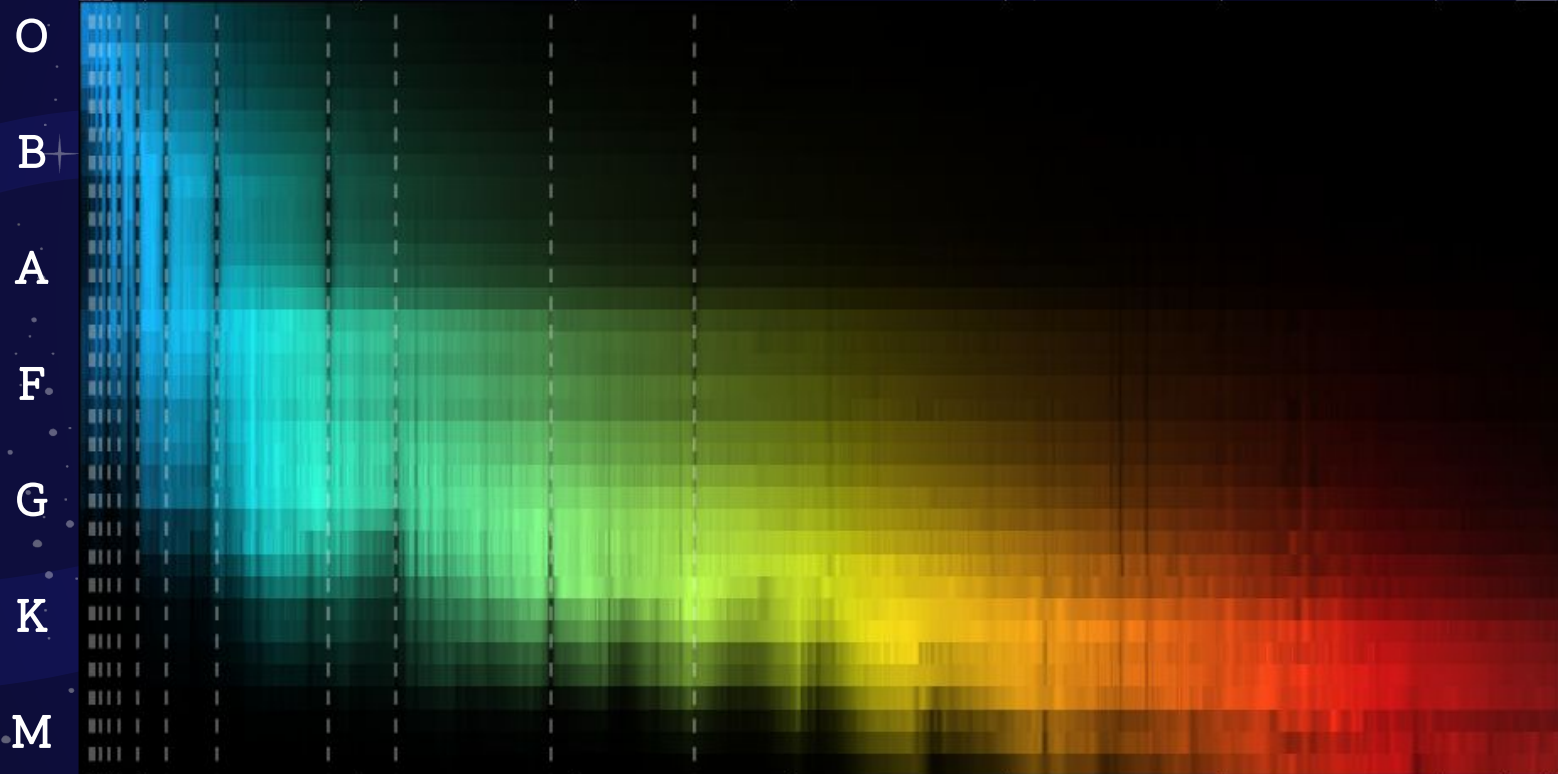
More of the Science

Spectrography

- ★ Different elements absorb and emit different wavelengths of light
- ★ Therefore we can determine presence of elements by looking for absorption lines (dark lines) and emission lines (bright lines)



Spectrography

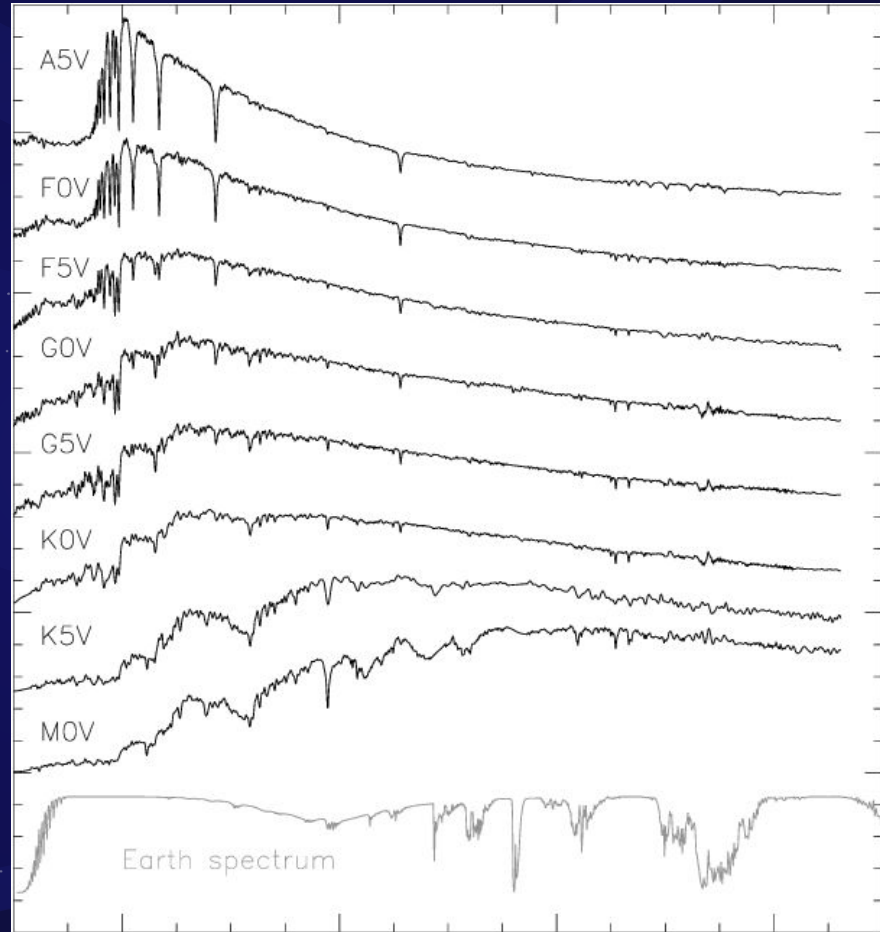


A bunch of different stellar spectra, ranging from Type O to Type M.

The different elements in these stars are causing lots of little lines.

Spectrographs

- ★ These spectrographs can also be represented as graphs like these.
- ★ Those absorptions and emission lines now show up as sharp spikes or dips on the graph.
- ★ The surface temperature of the star can be determined from these graphs by looking at highest peak.



Spectral Types

Normal spectral types range from blue hypergiants in Type O to brown dwarfs in Y (which emit almost no light).

The Normal Ones



But there are also abnormal spectral types, such as white dwarfs, relatively bright but very small stars

Spectral Types

The Itsy Bitsy Ones

Usually these are the core of a larger star that has blown off its outer layers in a supernova

White Dwarf

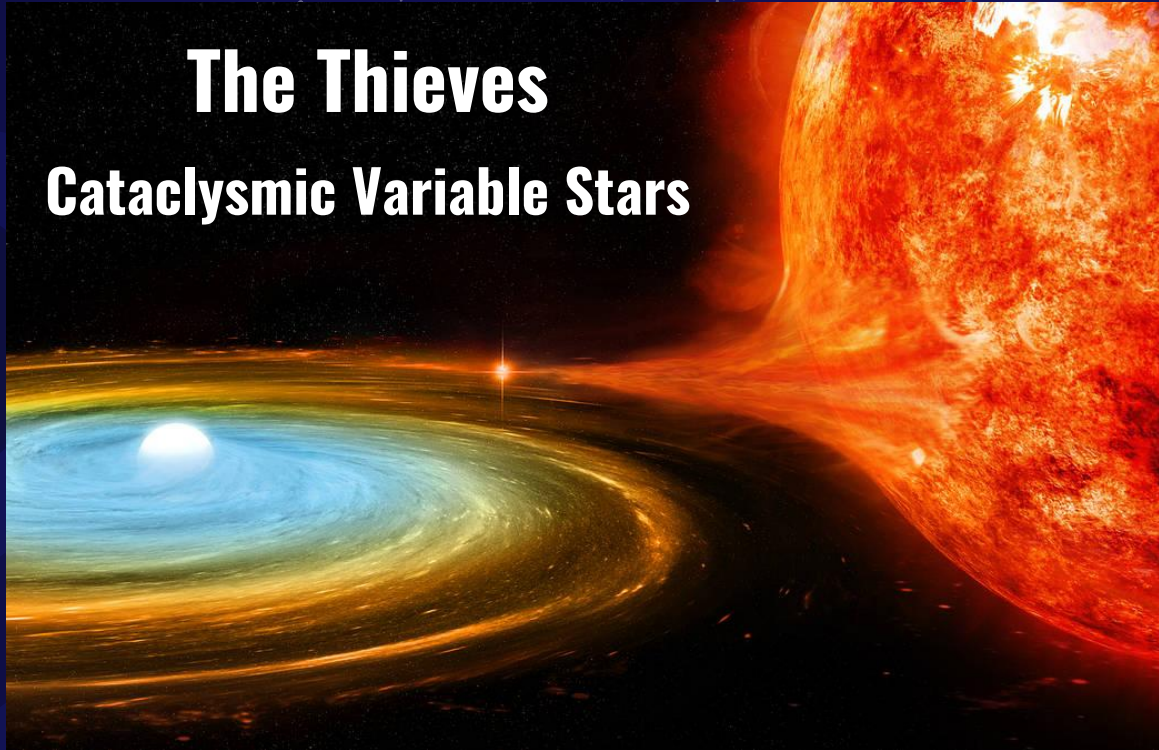
**Our
Sun**



Spectral Types

Cataclysmic variable stars are a type of star that fluctuates in brightness over time. These stars only appear in binary systems

The Thieves **Cataclysmic Variable Stars**

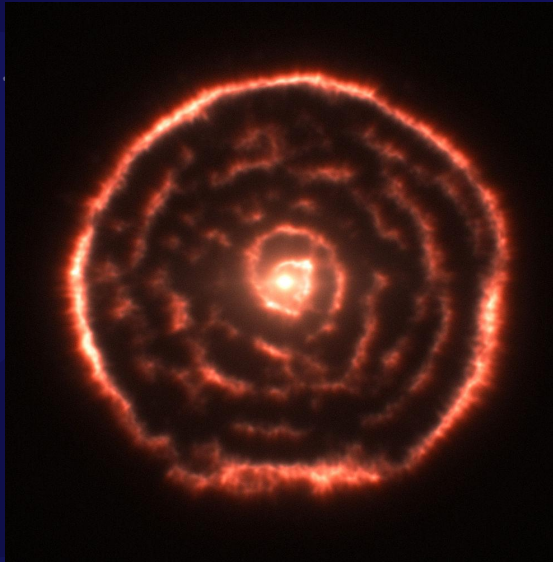


CV stars siphon material off of their binary companion, periodically building up layers and then blowing them off in a bright flash

Spectral Types

Carbon stars are the faint remnants of late stage (dying) stars that didn't have enough mass to supernova, so once fusion stops in the core, the outer layers gradually drift off

The Dying Giants



Carbon Stars

