Stellar Classification of stars in Celestial Space

Akhil Chowdary Maddipatla, Deenadayalan Dasarathan, Zijie Wei

Master of Science in Data Science, Khoury College of Computer Sciences,
Northeastern University, Boston, Massachusetts

Abstract

In our project we use the star data set which contains several features of stars which help us prove the type of stars in space and the type of certain graph in celestial space. We use labelled sources to train an optimized random forest classifier algorithm.

We apply this machine learning algorithm on stars which have existing observations and try to classify them based on the star type which is based on various factors. Here we use K-means and Hierarchical Clustering.

The reason why we use K-means and Hierarchical Clustering is because they are relatively simple to implement and generalize to clusters of different shapes and sizes which is especially the case with our dataset. We primarily use Hierarchical clustering helps us group similar kind of stars together.

Also, we have used the dataset which was publicly outsourced in Kaggle. In the dataset, we remove that target from our data set so that it aligns with our Unsupervised learning.

This dataset includes several details in Astrophysics. They are:

Stefan-Boltzmann's law of Black body radiation, which is find the luminosity of star, Wienn's Displacement law which finds the surface temperature of a star using wavelength, Absolute magnitude relation, Radius of the star using parallax.

https://www.kaggle.com/deepu1109/star-dataset

There are 6 features of stars which are absolute temperature, relative luminosity, relative radius, absolute magnitude, star color, spectral class.

- Absolute temperature shows the surface temperature of stars, and the temperature's unit is K.
- Relative luminosity shows the luminosity of those stars calculated with respect to sun.
- Relative radius consists of the radius of stars calculated with respect to sun.
- Absolute magnitude shows the absolute visual magnitude, which is Mv, of those stars. Star color shows the colors of each star after spectral analysis. These features will help us determine the final classifications.
- Lo = 3.828 x 10^26 Watts (Avg Luminosity of Sun)
- Ro = 6.9551 x 10^8 m (Avg Radius of Sun)

Star type shows the type of stars. The types include brown dwarf, red dwarf, white dwarf, main sequence, supergiant, and hypergiant. Each type is represented by a number from 0-5 respectively. Spectral class shows the spectral classes of each star. (This is the target which will be removed)

For EDA Unsupervised Machine Learning, we plot each feature against the remaining feature to see which star types are easily separable from each other. Here we also refer to the process of discovering patterns with the star dataset.