PROJECT PROPOSAL

Sara Gholamhoseinian

TITLE OF RESEARCH

Stars type classification by data analysis

OVERVIEW

stellar classification, scheme for assigning stars to types according to their temperatures as estimated from their spectra. One of the most significant classifications in astronomy, is the Hertzsprung-Russell diagram. The Hertzsprung-Russell diagram shows the relationship between a star's temperature and its luminosity. It is also often called the H-R diagram or colour-magnitude diagram. It is a very useful graph because it can be used to chart the life cycle of a star. (Fig. 1)

stars' surface temperatures in Kelvin is on the x-axis (horizontal axis) and stars' luminosity (or absolute magnitude) along the y-axis (vertical axis). the x-axis of the H-R diagram can use different data. It might show the star's temperature, its spectral class (OBAFGKM), or its colour. All these types of data show the same relationship with a star's luminosity.

OBJECTIVES

- Prove that stars follow the H-R diagram, so that we can classify stars by plotting its features based on that graph.
- Investigate the relation between different features of stars and compare them with observational results
- analysing the correlation between characteristics of stars

- predict stars classes (Brown Dwarf, Red dwarf, White dwarf, Main Sequence, Supergiant, or Hypergiant)
- Finding which statistical model performs better for this dataset.

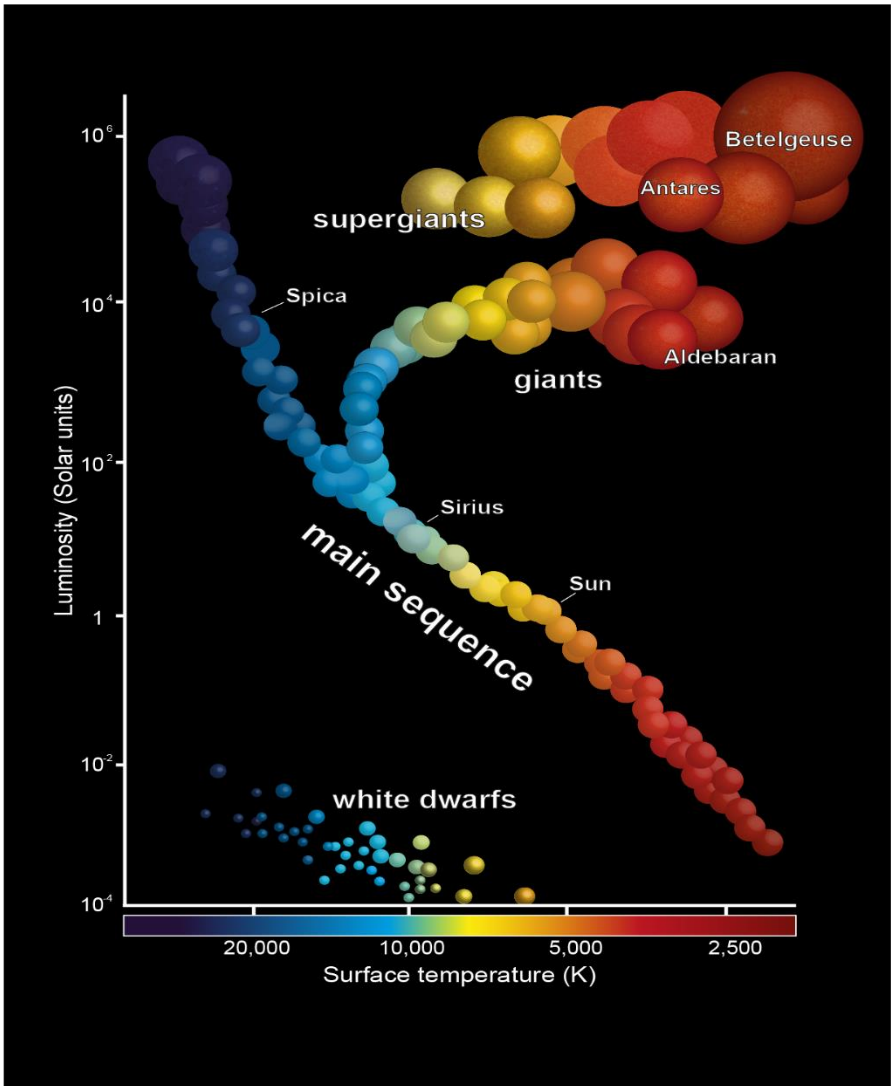

Fig1. The H-R diagram

DATA COLLECTION

This dataset has been created based on several equations in astrophysics. They are given below:

- Stefan-Boltzmann's law of Black body radiation (To find the luminosity of a star)
- Wienn's Displacement law (for finding surface temperature of a star using wavelength)
- Absolute magnitude relation
- Radius of a star using parallax .

I found this dataset on Kaggle and it consists of several features of stars such as :

- Absolute Temperature (in K)
- Relative Luminosity (L/Lo)
- Relative Radius (R/Ro)
- Absolute Magnitude (Mv)
- Star Color (white, Red, Blue, Yellow, Orange, etc.)
- Spectral Class (O,B,A,F,G,K,,M)
- Star Type (Red Dwarf, Brown Dwarf, White Dwarf, Main Sequence, Supergiants, Hypergiants)
- Lo = 3.828 x 10^26 Watts (Average Luminosity of Sun)
- Ro = 6.9551 x 10^8 m (Average Radius of Sun)

DATA ANALIZING

in this project, I will take advantage of machine learning methods and split the dataset into train data and set data to process the data for training and start data visualization. Data visualization is a key aspect of ML. First, we need to know that most of the stars in the universe can be classified based on the H-R Diagram of stars shown previously.

After analysing the original tested data, if the predicted data is almost similar (at least 98% accurate) to that of the tested data then we can say that the classifier learned efficiently (In our case we got 100% accuracy with no overfitting nor underfitting).

Moreover, I will use some of the libraries in python such as NumPy, seaborn, matplotlib, pandas, and Sklearn. Also, we will calculate some of the important features of the dataset such as skewness, median, … to gain more information from the dataset. In the end, we will compare the model accuracy in different methods of ML.

POSSIBLE ISSUES

This dataset contains information about only 240 stars, probably it will be enough for our project but it might have less generalization compared to the bigger datasets. So, I found another dataset that is more advanced. if I realize that the primary dataset is not enough reliable and it can't lead me to the desirable result, I would replace it with the second one.

RESOURCES

1. Stellar classification – Wikipedia
2. Stars spectroscopy – science direct
3. The physics of stars – A. C. Phillips
4. Star type classification - Kaggle
5. Stellar classification dataset - Kaggle