Our project focused on the classification of stars through the use of machine learning. In this project, we look to develop models that can accurately categorise stars into their respective classes, using a variety of data analysis techniques and tools. After forming our group and gathering together, we searched for various datasets across different categories

**Why we picked the stellar dataset?**

1. Abundance of data: The stellar dataset we looked at was very large and provided us a wealth of data for our machine learning algorithms to work with. This was particularly useful when working with deep learning models, which as you know require large amounts of data to train effectively.
2. High-quality data: The Stellar dataset was well-organized and carefully curated, with data points that had been meticulously observed and recorded. This helped ensure that the data was of the highest quality and lead to more accurate machine learning models.
3. Rich features space: The Stellar dataset contained a wide range of features that we could use to train machine learning models. These included properties of stars themselves, such as their temperature, luminosity, and chemical composition, but their were other features such as environmental factors such as the presence of other stars and interstellar gas and dust that we didn’t use.
4. Scientific interest: The Stellar dataset was of great interest to us as we were interested in developing machine learning models that can be used to gain new insights into the properties and behaviours of stars.

Overall, The Stellar dataset provided a rich source of data for our machine learning project, with the potential to lead to new scientific discoveries and advances in our understanding of the stellar bodies.

**AIM – What?**

In our study, we plan to utilize the MK (Morgan Keenan) classification system to categorize stars based on their chromaticity (Hue) and size using the spectral dataset. We focused on categorizing stars into their main spectral types using the Absolute Magnitude (Absolute magnitude is a measure of the intrinsic brightness of a celestial object) and the B-V Color Index (The B-V color index is a measure of the colour of a celestial object, such as a star. It is defined as the difference between the apparent magnitudes of the object in two different wavelength bands: the blue ("B") band and the visual ("V") band) within our dataset. This approach we felt provided a useful way to understand the properties of stars within our dataset and to classify them based on their spectral characteristics, which can be important for understanding their evolution and behaviours.

Machine learning will be used to automate and improve the process of classifying stars using the Morgan-Keenan (MK) classification system. The traditional method of classifying stars based on their spectra requires human experts to visually analyze the absorption lines in the spectra and make subjective judgments about the star's spectral type.

We aim to use Machine learning algorithms to be trained on our large datasets of spectral data and use the corresponding MK classifications to learn the patterns and relationships between spectral features and spectral type. These algorithms will then be used to classify new spectral data with a high degree of accuracy and consistency, while also potentially uncovering previously unknown patterns or correlations.

Now that you understand the motivations and purpose of our study, we can discuss how we achieve our aim. ………

To ensure that only relevant data is analysed, we implemented an ETL process for this project. This process was implemented using Jupyter Notebook, SQLAlchemy, and Pandas DataFrame to help with the extraction, transformation, and loading of data. With this process, we can quickly and easily access the necessary data without the added complexity of different databases.

The first step is to collect spectral data on the stars we are interested in studying. We will then analyse the data to identify the absorption lines present in the spectra, which will allow us to determine the star's spectral type using the MK classification system.

To further refine our classification, we will also use the star's absolute magnitude and B-V color index. By comparing these values to established relationships between spectral type and these parameters, we can better categorize each star and gain insights into their chromaticity and size.

We will focus on categorizing stars into the main spectral types, ranging from O to M, as this provides a useful framework for understanding their properties and behavior. Through this approach, we hope to gain a deeper understanding of the stars in our dataset and contribute to the broader field of astronomy.

Bottom of Form