

## Populärvetenskaplig beskrivning

Contrary to the popular belief that most stars are singles, around half of the stars we see in the galaxy are actually found in pairs called binary systems or simply "binaries". Their binary nature can be discovered or inferred in many different ways, such as through eclipses that occur when one star of the pair passes in front of the other, or through the characteristic features and behavior of their combined spectrum. Moreover, binaries play a major role in astrophysics. They offer scientists an insight into crucial stellar processes as well as enable accurate measurements of fundamental stellar parameters such as mass and radius through the gravitational interaction between the two components of the system.

In recent years, stellar surveys have increased exponentially in complexity and amount of stars observed, and while binaries have been shown to be abundant they are often missed in observational data. The reason for this is that in order to reveal their true nature, binary stars require a special handling besides that given by traditional methods for the analysis of stellar data. This can, in most cases, be quite time consuming. However, new approaches for discovery and characterization of binary stars have been made possible by advances in the field of machine learning and the increase of computational power. Machine learning is the name given to a set of algorithms and statistical tools used by computers to extract information from data by recognizing patterns without being explicitly programmed to do so. With it, it is possible to not only examine and study the large amounts of new data gathered by stellar surveys, but also "revisit" older data-sets in order to extract insights and patterns that were overlooked in the past.

In this project we will try to address the discovery efficiency of binary stars in an archetypical spectroscopic survey when using machine learning algorithms. By generating different types of spectra ourselves, we create a mock spectroscopic survey data-set for which the distribution of stellar parameters and the amount of binary stars are known. Unlike in real surveys, by using self-generated data the nature of each star is known beforehand. This allows us to evaluate a series of machine learning algorithms with respect to its own input parameters and the ranges of stellar parameters present in the generated data. With this evaluation, we want to determine and constrain the efficiency and limits of the used method regarding their efficiency discovering and detecting binary stars.

Our aim is to generate an automated method capable of maximizing the recovery and detection of binary systems from real spectroscopic data while being scalable and applicable to future surveys. To achieve such a goal, we combined two well-known and readily available machine learning algorithms for the automatic analysis of spectroscopic data and the retrieval of binary stars from it. One algorithm is called t-SNE, which is used to project the data onto a plane, grouping objects that are similar in clusters and separating those that are dissimilar. The groups of data-points representing spectra created by t-SNE are then automatically recovered regardless of their morphology by the second machine learning algorithm we use, which is called DBSCAN.

The combination of t-SNE and DBSCAN, whose individual implementation has been carefully chosen to minimize the computation time, allowed us to obtain results that are easy to implement and understand. Results from our study are promising, showing a mean recovery of 57%, averaged over all the 360 simulations we carried. We find that the presence of moderate noise levels in the studied spectra can help improving the detection of spectroscopic binaries, as it can smear out information from it that might throw off the machine learning analysis. Furthermore, we show that bluer spectral regions (between 450 and 650 nm) are better suited than those in redder parts of the spectrum (between 650 and 900 nm) due to the increased amount of information in the form of spectral lines present in the analyzed spectroscopic data. In the end, we provide a table of stellar parameters for binary stars that were contained in our synthetic

---

sample and which were successfully identified in more than 90% of our simulations and which can serve as a guide for future implementations of our method.