Big Data Algorithms (CS 462/562)

This course provides a rigorous introduction to algorithms and techniques for processing very large data sets, including both offline and streaming data. The emphasis will be on developing a clear understanding of the **theoretical underpinnings** of algorithms for **data mining** (*e.g.*, similarity detection, clustering) and **machine learning** (*e.g.*, supervised and unsupervised learning models, regression, support vector machines). Python-based libraries and software like *Pandas*, *Scikit-learn* and *Pyspark* will be used to develop applications for large datasets that are based on such algorithms.

> **ℹ Note**
>
> This is a hands-on course: becoming proficient in the use of standard tools and libraries in the Python programming ecosystem will be emphasized alongside theoretical understanding. **Every student must have their own laptop that they bring to the class sessions!**

SYLLABUS

- Fall 2024 Syllabus