A Comparison of Common CS topics

Data Science (DS) is a multi-disciplinary approach towards understanding a dataset (typically on large amounts of data, i.e., Big Data). It combines the understanding of many disciplines such as business, analysis skills from statistics and mathematics, programming skills, as well as the additional knowledge of the subject dataset. These different skills result in Data Science being a multi-dimensional discipline requiring lots of spread-out knowledge.

Machine Learning (ML) on the other hand is a smaller, more concise discipline. It can be categorized as 4 different categorical learning styles: Supervised, Unsupervised, Semi-supervised, and Reinforcement. Each style of learning takes a different approach for the machine to learn from the given dataset. Additionally, there are different outputs based on the approach used. These algorithms are collectively known as Machine Learning Algorithms and are used by Computers to gain insights into data, typically returning relations between variables.

Data Mining (DM) is commonly confused with ML because it also looks at data and is used to extract knowledge. The big difference between Data Mining and ML is that Data Mining looks for patterns within datasets. It finds previously unknown information and looks for implicit relationships or patterns that may exist within the given dataset. These processes engineer and systematically break down the dataset to gain insights on the data.

Artificial Intelligence (AI) is a hot topic right now with the recent surge of popularity of Chat-GPT. Chat-GPT falls under the branch of Generative AI which is a subset of DL, ML, and AI itself. Artificial Intelligence is the largest of the grouping, as it combines techniques from DL, ML, and more whilst using more advanced techniques understood from social sciences such as Psychology. Using such a combination of disciplines allows for Machines to simulate learning as a human, thus the name Artificial Intelligence.

Deep Learning (DL) is the process to learn relations amongst 2+ variables. Additionally, DL analyzes the reasons and understandings that explain why there is a relation amongst the variables. This allows for comprehension of the relationship between Input and Output as well as the parameters that affect such a relationship.

1. Y. Demchenko et al. “EDISON Data Science Framework: A Foundation for Building Data Science Profession for Research and Industry” 2016
2. I. K. Nti, et al. “A mini-review of machine learning in big data analytics: Applications, challenges, and prospects” 2022
3. L. Xu, et al. “Using deep learning models to predict student performance in introductory computer programming courses”2014
4. Y. -H. B. Chiang, et al. “Using deep learning models to predict student performance in introductory computer programming courses” 2022
5. J. Harika, et al. “A Review on Artificial Intelligence with Deep Human Reasoning” 2022
6. Y. Huang et al. “Identification of data mining research frontier based on conference papers” 2021
7. W. J. Zhang, et al. “On Definition of Deep Learning” 2018