Vinoothna Bavireddy

Terminologies of Data Science: Basic Understanding

Big Data: The term big data is coined to refer to data that is complex, large and is difficult to process using traditional methods. The definition of big data is defined through the three V's articulated by analyst Doug Laney which are:
1) Volume- Using Big data one can process enormous volumes of low-density unstructured data even those with unknown values. The data size can range from tens of terabytes to thousands of petabytes.
2) Velocity- Big data is capable of acting on data in real-time which makes it extremely valuable for any industry as data is received and acted upon at a very fast-rate
3) Variety- Big data supports not just structured data but unstructured or semi-structured data types allowing for analysts to derive meaning from varied different data types those including text, audio and video.

Hadoop is a software framework used for storing large amounts of data and computing big data fast. It is flexible in terms of data storage and unlike traditional databases the unstructured data need not be pre-processed before storage. This open source framework is cost-effective and can support concurrent tasks at the same time.

Data Science: On a basic level data science refers to making meaningful insights from data. These insights are driven by analyzing data through tools such as SQL, Python or R which refers to the subcategory of data science called Data Analysis. The other subcategories include Experimentation (running an experiment by keeping the data clean and analyzing the data), Machine learning (Using machines to understand, process or replicate a system), artificial Intelligence (using human interaction as a means to provide insights).

"**Data mining** is defined as a process used to extract usable data from a larger set of any raw data." Data Mining helps accelerate to derive insights from the data to identify patterns in huge datasets, predicting likely outcomes, creating decision oriented information, managing and eliminating risks.

References:

1.Oracle. (n.d.). What is big data? Oracle.com. Retrieved from :
https://www.oracle.com/big-data/what-is-big-data.html
2.SAS, Hadoop what it is and Why it matters. Retrieved from:
https://www.sas.com/en_us/insights/big-data/hadoop.html
3.Thinkful. (n.d.). What is data science? https://www.thinkful.com/blog/what-is-data-science/
4.Data mining. (n.d.). The Economic Times.
https://economictimes.indiatimes.com/definition/data-mining