SAMWEL JOHN PETRO

Qn. why is Knowledge Discovery in Database synonymously called data mining given that data mining is a step of Knowledge Discovery in Database

The term knowledge discovery in databases is knowns as or synonymously called data mining due to the following reasons.

1. It is a Core step of knowledge discovery: knowledge discovery also people call it as data mining, because data mining is a main step in which data are extracted from larger data set and being used to solve problems or make decisions. Other steps prepare, clean and interpret while data mining is where the actual information or required information are being extracted from larger dataset.
2. It focuses on result or practice: data mining focus on result or success or outcome of the whole process of knowledge discovery, hence making the entire process of knowledge discovery to be associated with this step.
3. Simplification or easy of communication: data mining is shorter and ease word to use in communication compared to knowledge discovery in databases thus making it to be used interchangeably meaning the same thing
4. Historical and practical usage: most of people after knowledge discovery being introduced, they started to use data mining as the word to mean the whole process, also most of algorithms used in this process are known as mining algorithm hence making data mining being used interchangeably with knowledge discovery.
5. Popularity: most of people uses data mining to means knowledge discovery in databases hence making it widely adopted to others and being used interchangeably.

Qn. What are the differences between data wrangling, scanning and filtering in data mining?

Data wrangling	Data scanning	Data filtering
i. refers to the overall process of cleaning, transforming and organizing raw data into a usable form for data mining.	i. refers to the process of sequentially reading or examining data records to understand their characteristics or locate required information. .	i. refers to the process of selecting some or subset of data by removing unwanted records based on a specific condition.
ii. it involves preprocessing of data.	ii. it involves inspection of data.	ii. involves reducing data size and noise.