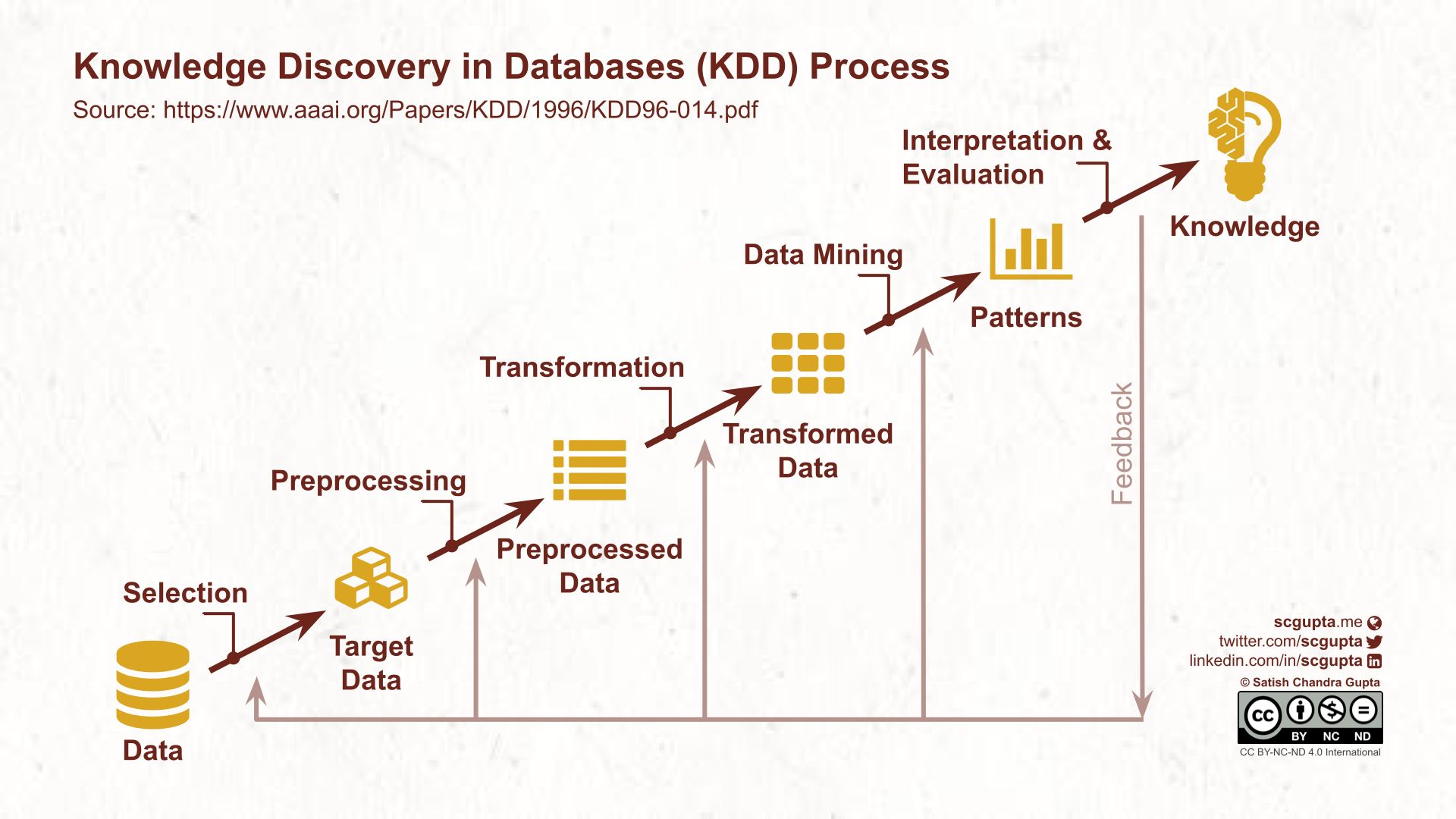
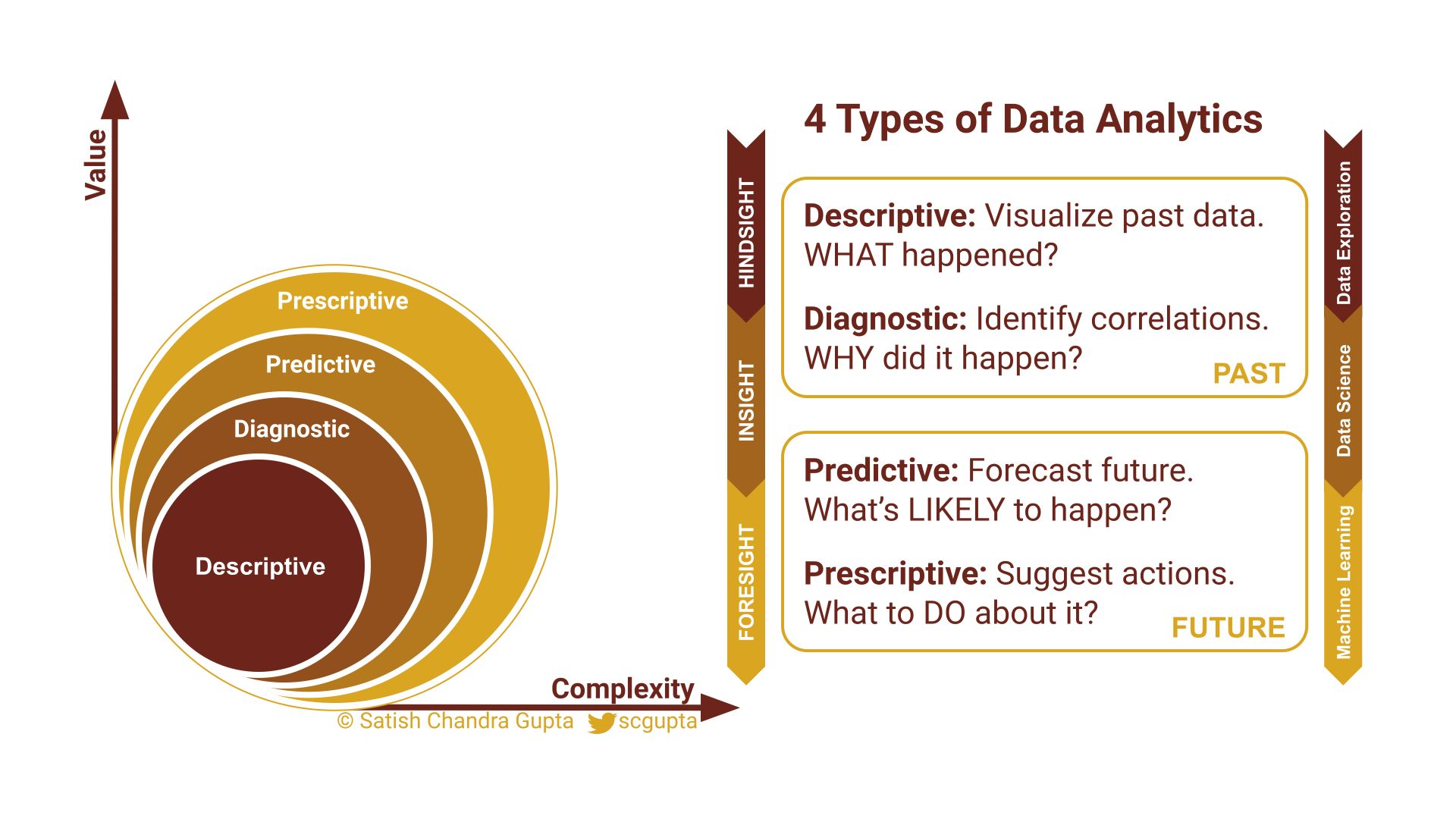


Do you ever wonder what the difference is between Data Science (DS) and Machine Learning (ML)?  
  
Both require building statistical models.  
  
My differentiation is biased by a software engineer's worldview:  
  
ML is automating DS to repeatedly perform it at scale because DS always has a human in the loop.  
  
👉 Artificial intelligence (AI):  
Intelligence demonstrated by machines by performing like humans.  
  
👉 Machine learning (ML):  
Use of statistical models to make predictions by recognizing patterns in the data, and without requiring explicit instructions.  
  
👉 Deep learning (DL):  
Machine learning methods using deep neural networks.  
  
👉 Data science (DS):  
Use of statistical techniques to extract knowledge and insights from structured and unstructured data.  
  
Recently I read an "oversimplified" definition of the differences that felt quite right:  
🔸 Data Science produces insights  
🔸 Machine Learning produces predictions  
🔸 Artificial Intelligence produces actions  
  
Deep Learning is ML with Deep Neural Nets.  
(Link to the article is in the first comment)  
  
How do you distinguish between these fields? Should we really care?  
  
Click [#ML4Devs](https://www.linkedin.com/feed/hashtag/?keywords=ml4devs&highlightedUpdateUrns=urn%3Ali%3Aactivity%3A6769841480916987904) and follow for more content on Machine Learning for Developers.



MLOps evolution is fascinating.  
  
Information Retrieval and Data Mining have been around for 3 decades. One of the first processes was KDD (link to the original research paper in comments).  
  
"Knowledge Discovery in Databases (KDD) is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data."  
  
Steps:  
1. Develop application understanding and define goals  
2. Create a target dataset  
3. Clean and preprocess data  
4. Transform the data and find useful features, use Dimensionality Reduction to reduce variables  
5. Match goals to a suitable data mining problem: classification, regression, clustering, summarization, etc.  
6. Choose data mining algorithm(s)  
7. Perform data mining  
8. Interpret mined patterns  
9. Consolidate discovered knowledge  
  
Isn't the Data Science steps the same as above?



Gradually expanding Data Analytics teams to do Machine Learning is an attractive approach.  
  
Especially for companies dealing with mostly tabular data.  
  
There are 4 stages of data analytics (each more complex & valuable than the previous):  
  
1. Descriptive: WHAT happened?  
  
Slice, dice, and visualize the data. Organize information. Aggregate various dimensions. Report what, when, where, how many.  
  
2. Diagnostic: WHY did it happen?  
  
Drill down into details to identify correlations and root causes. Identify outliers, isolate patterns and uncover relationships. Techniques such as regression and time series analysis are used to understand cause and effects.  
  
3. Predictive: What's is LIKELY to happen?  
  
Focus shifts from past to future. Use discovered patterns and correlations to forecast probabilities of potential future outcomes. Various quantitative analysis and statistical models are used. Pretty much like Machine Learning.  
  
4. Prescriptive: What to DO about it? HOW to make it happen?  
  
Actionable analytics. Recommend actions for optimizing outcome. Search through "what-if" scenarios on various levers available. Use predicted probabilities and correlations to devise the best course of action.  
  
What kind of analytics are you currently doing?

<https://www.linkedin.com/in/scgupta/>