



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Lecture 18.1 - Machine Learning for Streaming Data Analysis

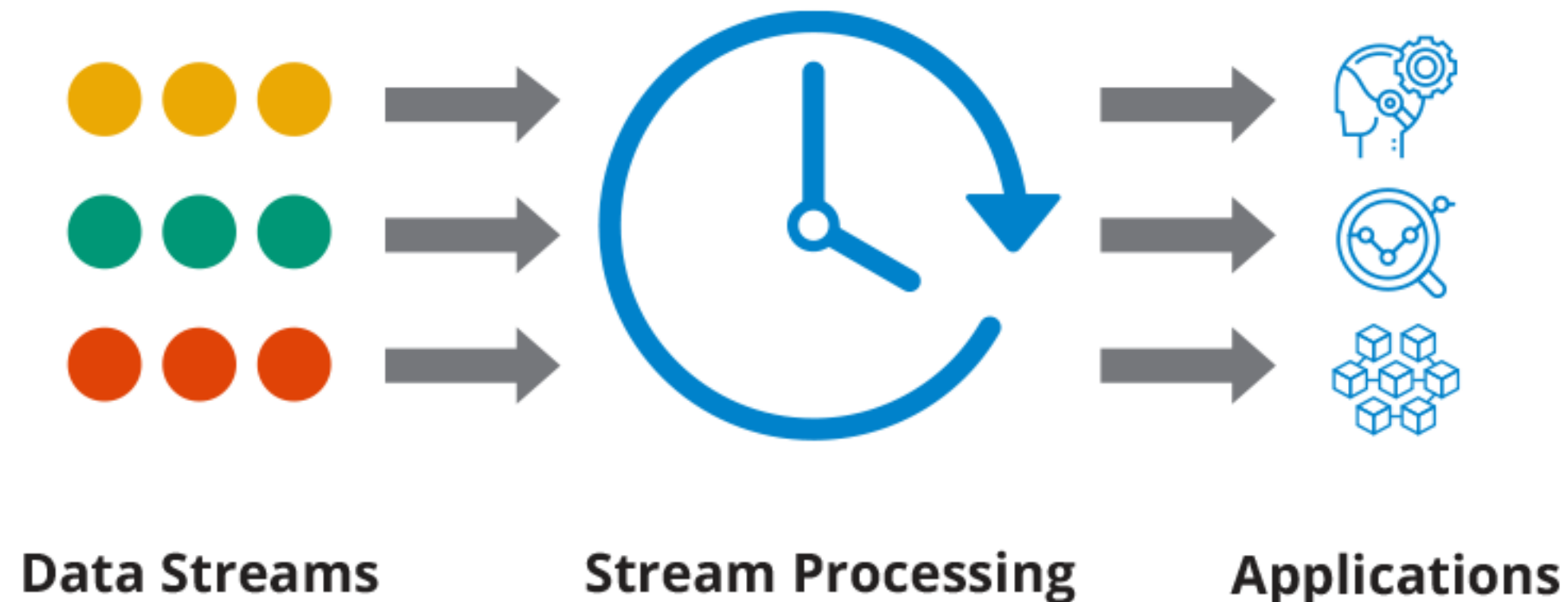


The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

Streaming Data

Streaming data is generated continuously by various data sources, typically sending the data records simultaneously.

- It is the continuous flow of data generated by different sources, in various formats and volumes.
- It is generated by applications from networking devices, and server log files, to website activity, banking transactions, and location data.



Streaming Data Application

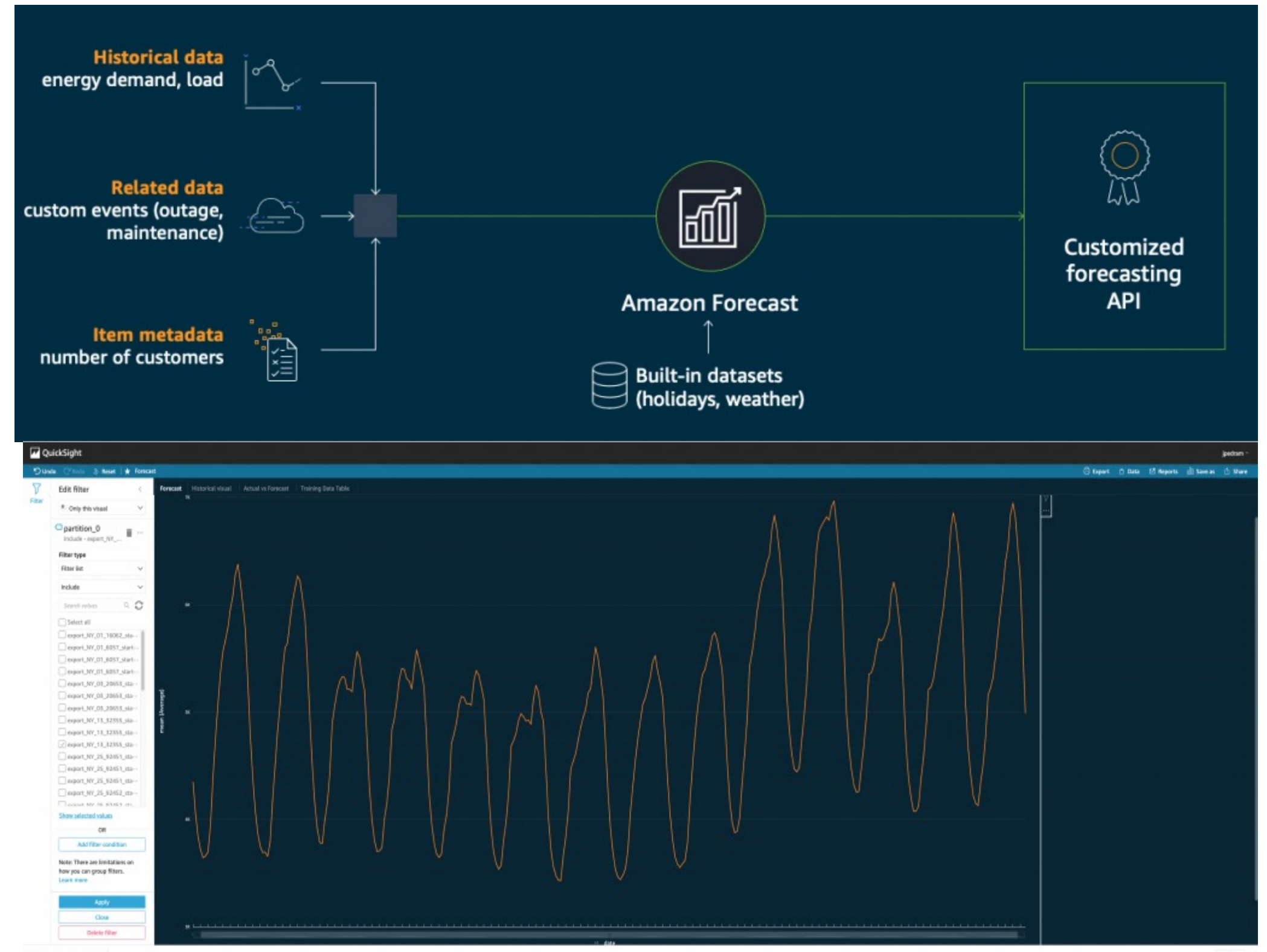
Real-time stock trades

Power load forecasting

Social media data analysis

Multiplayer game interactions

Ride-sharing application

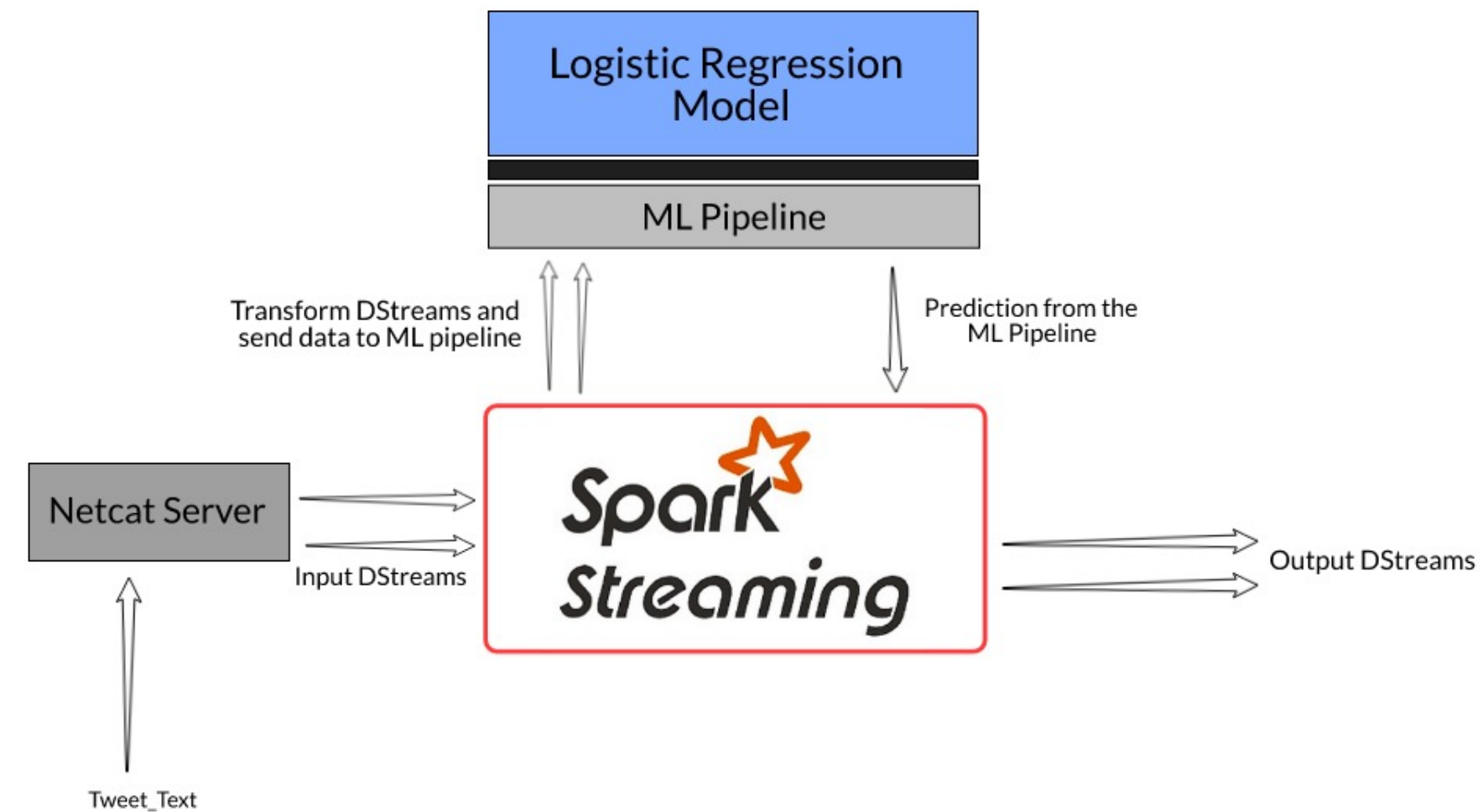


Source: <https://tdwi.org/articles/2020/07/13/arch-all-new-approach-to-streaming-analytics-million-data-sources.aspx>

Machine Learning for Streaming Data

Machine learning based streaming data analysis employed supervised learning models, especially classifiers to address the problem of changes to the underlying data distribution over time.

In general, it is to overcome one specific challenge: developing methods that maintain an accurate decision model with the ability to learn and forget concepts incrementally.



Tasks

Data stream clustering

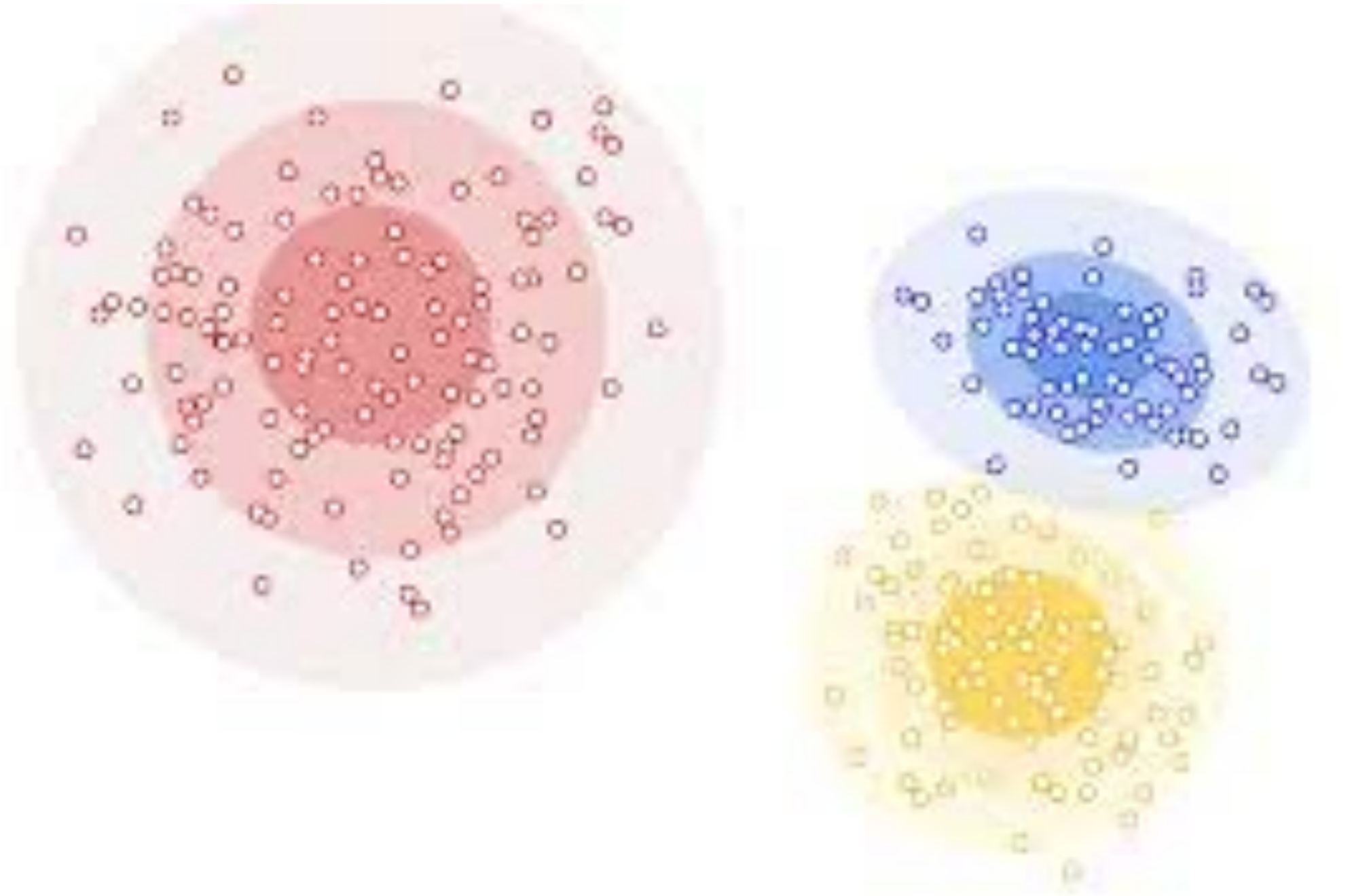
Pattern mining

Anomaly detection

Multi-output learning

Novel class detection

... ..





DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Thank You