Juan Póez - 20231020028

T10: What is a Feature Store and what is Data Lineage?

Feature engineering is a crucial process in machine learning that involves creating new features from existing data to improve model performance. This process is complex and labor-intensive, but it is critical to achieving better business outcomes.

Feature store is a solution to the dispersion of features across different locations. It acts as a centralized repository that allows data scientists to efficiently find, share, and reuse features. It also offers a data transformation service, making it easier to prepare data for use in machine learning models.

Data linage is another important concept relate to feature management. It invoves tracing the journey of data over time, from its source to its final destination. This allows you to understand how the data was transformed, detect errors, and ensure data quality. In summary, the text highlights the importance of feature engineering and the need for tools that facilitate this process.

Feature store and data lineage are two key elements to achieve efficient and reliable data management in machine learninang projects. Data lineage tools provide a record of data throughout its lifecycle, including source information and any data transformations that have been applied during any ETL or ELT processes.

This type of documentation enables users to observe and trace different touch points along the data journey, allowing organizations to validate for accuracy and consistency. This is a critical capability to ensure data quality within an organization. It is commonly used to gain context about historical processes as well as trace errors back to the root cause.