Luis Anton P. Imperial
BCS32

S-CSIS311
Introduction to Machine Learning (Lab)

Wed, Aug 21, 2024
Learning Task 1

The Nature of Data Processing

Luis Imperial – Week 1 – Module 1: Fundamental Machine Learning.

Data processing is a key part of creating the artificial intelligence models we take for granted today. It is a process that distills raw input taken from vast quantities of end-users into usable and useful solutions in our daily workflow.

Concept Map
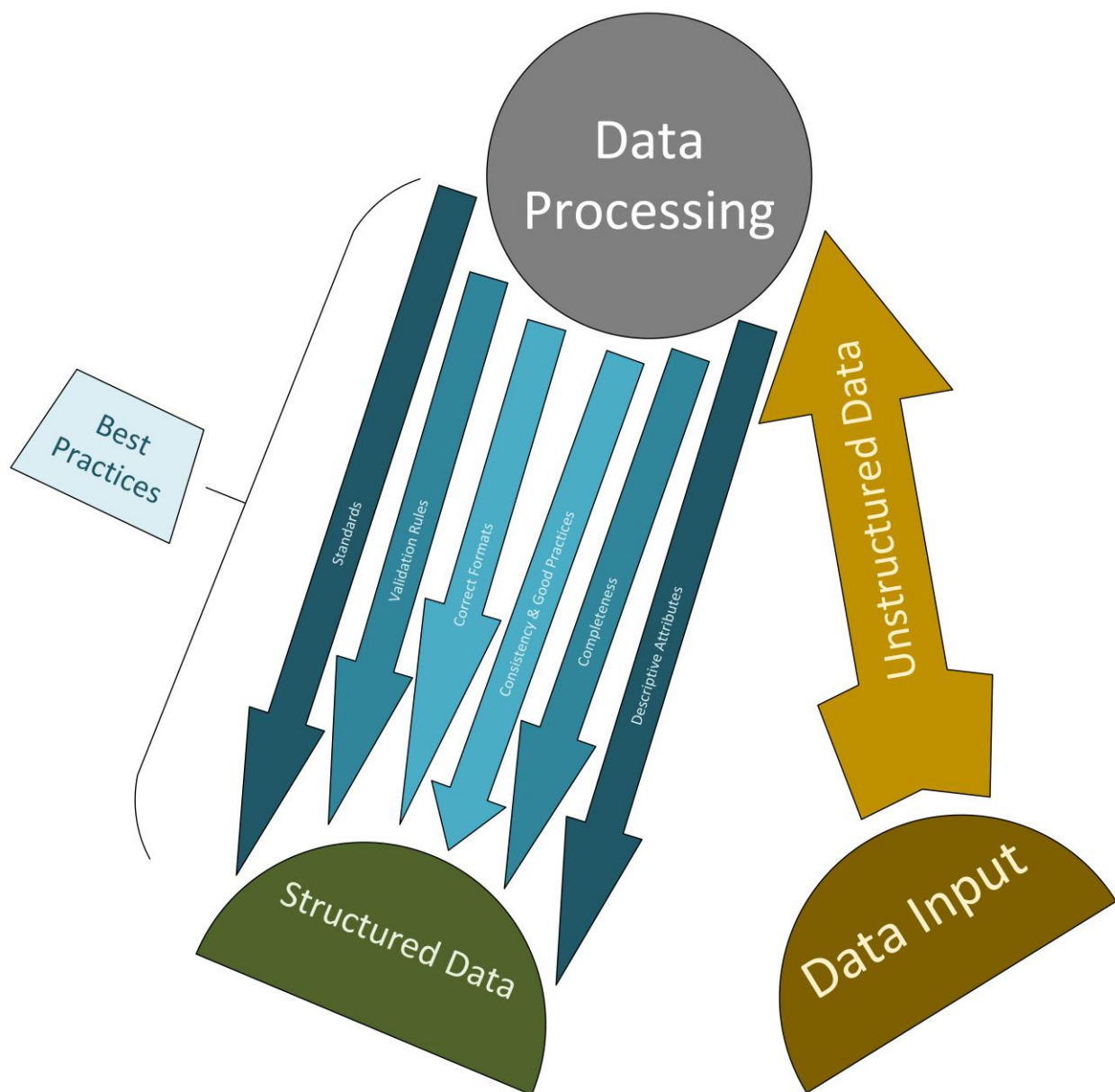
Discussion

Concept Map Discussion

The concept map represents **data** as primarily broken down into three main elements: **data input**, **data processing**, and **structured data**.

End-users of a solution, such as individuals and organizations, input data they have primarily as **unstructured data**. This then goes through the **data processing** activity, which is necessary for tools to analyze the inputs to provide the information requested from clients.

Data processing may include following a common set of standards, being approved under prepared validation rules, conversion into the correct formats, requiring consistency and following good practices, checking for completeness of sent entries, and classification into attributes that are described in detail.

Processed data may also be known as **structured data**, which can be absorbed more easily by machine learning algorithms and by business use-cases.

Additional Discussion

While this concept map focuses on the process of structuring raw (unstructured) data, it is important to know the several benefits and disadvantages of each type of data.

Unstructured data provides users the freedom to submit their inputs in any format, which may widen the scope of information to gather from. They may also be collected faster, as they can be quickly stored in databases and file systems.

Structured data is better suited to mass production of presented information, like dashboards of customer activity in each business cluster. This, of course, requires customer input to go through the process mentioned in the previous section.

References

- Barrameda, Rolando B. (August 21, 2024). "Learning Task." Computer Science Department, via Schoolbook for DLSU-D College/GS. *Dasmariñas, CAV: De La Salle University – Dasmariñas*.
- Steve, Chris. (April 19, 2018). "Top 10 Best Practices of Data Entry." Offshore India Data Entry. *Kakkanad, KL: Offshore India Data Entry*.

- Talend (n.d.). "Structured vs. Unstructured Data: A Complete Guide." Knowledge Center. *King of Prussia, PA: Talend, Inc.*