



CS 329P : Practical Machine Learning (2021 Fall)

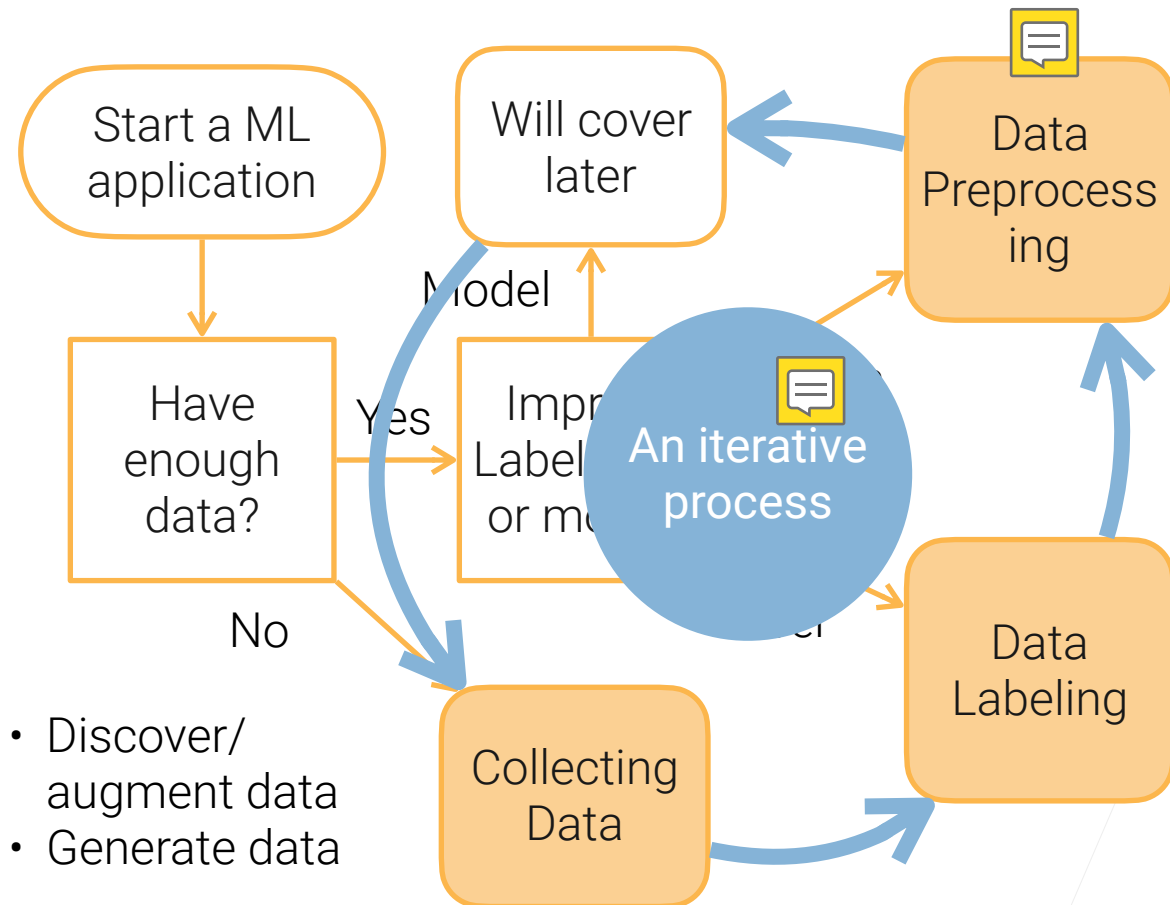
2.5 Data Part Summary



Qingqing Huang, Mu Li, Alex Smola

<https://c.d2l.ai/stanford-cs329p>

Flow chart



- Discover/ augment data
- Generate data

- Exploratory data analysis
- Data cleaning
- Data transformation
- Feature engineering

- Semi-supervised learning
- Label via crowdsourcing
- Weak supervision