



# Machine Learning for Business Processes

**Ashish Upadhyay**

PhD Student, Computing Science (AI), RGU Aberdeen  
Alumnus, BTech (Class of '19), IIT Naya Raipur

# Outline

- Business Processes
- Machine Learning
- Labelled Data
- Active Learning
- Closing Notes

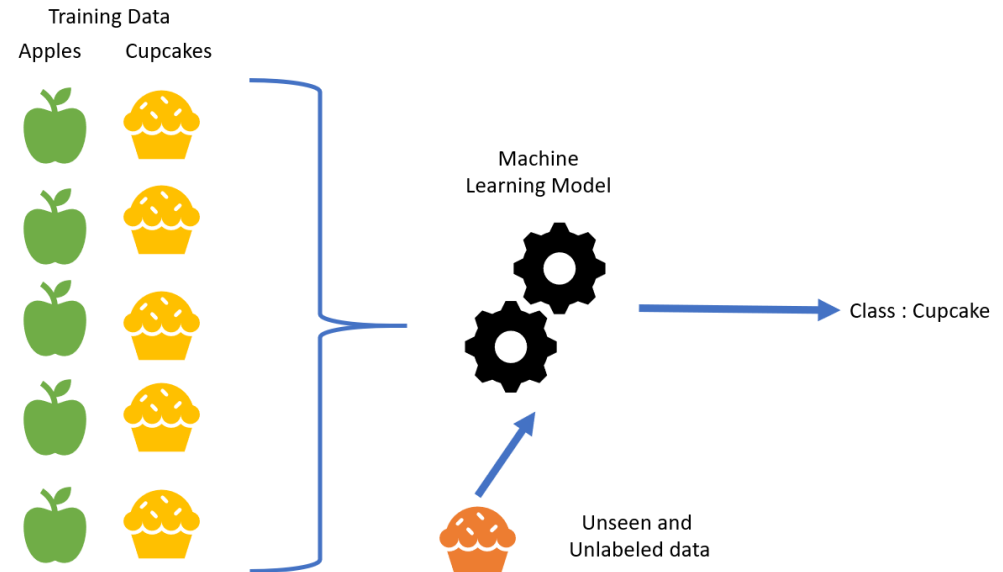
# Business Processes

- A small process of some big business
- Take an example:
  - Business: Regulatory Management
  - Process: Identify the source of regulatory documents
- Classifying a bunch of documents
  - A Machine Learning task



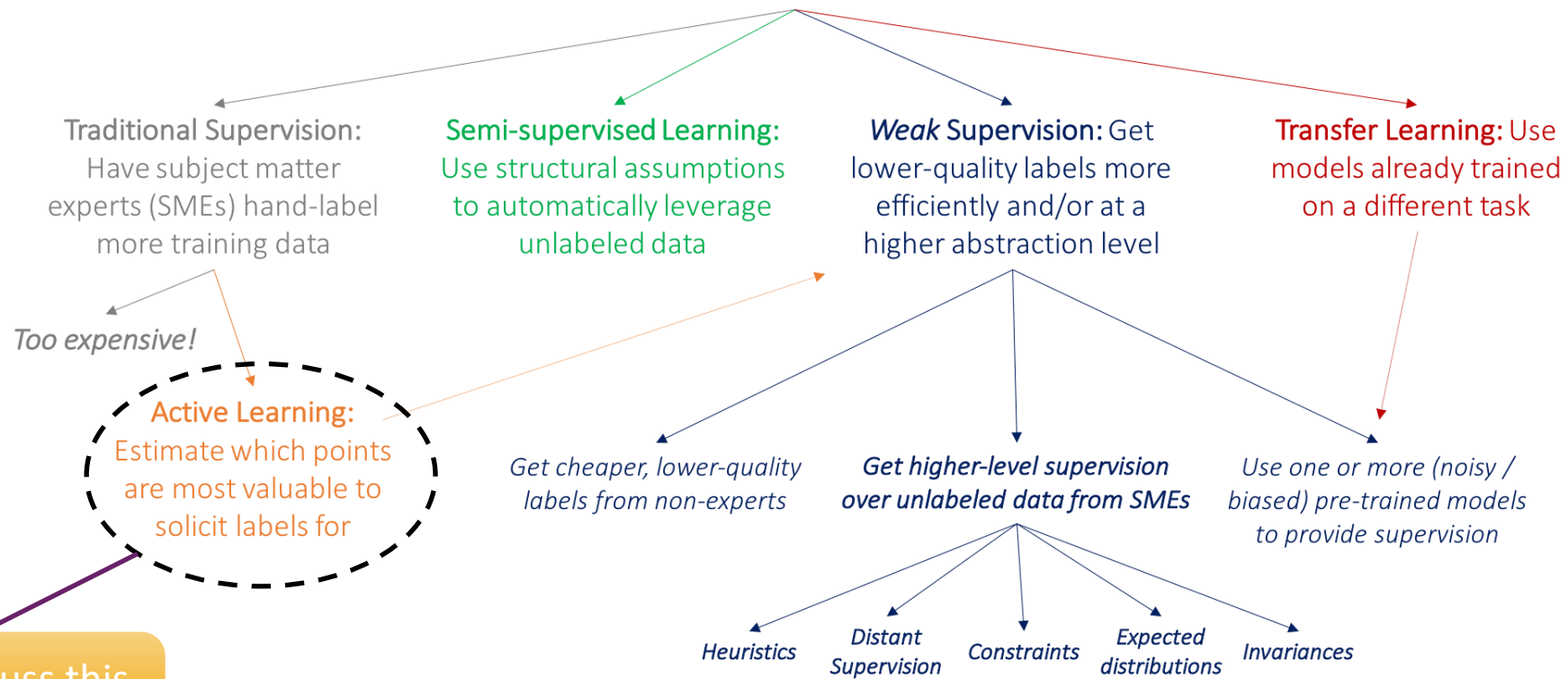
# Machine Learning

- Learning relation between input and output space
  - Using labelled data
- We need **Data, Labelled Data!**
  - ML algorithms are data hungry
- Getting labelled data
  - Expensive and time consuming
- Biggest question:
  - How do we get labelled data?



# Labelled Data

How to get more labeled training data?

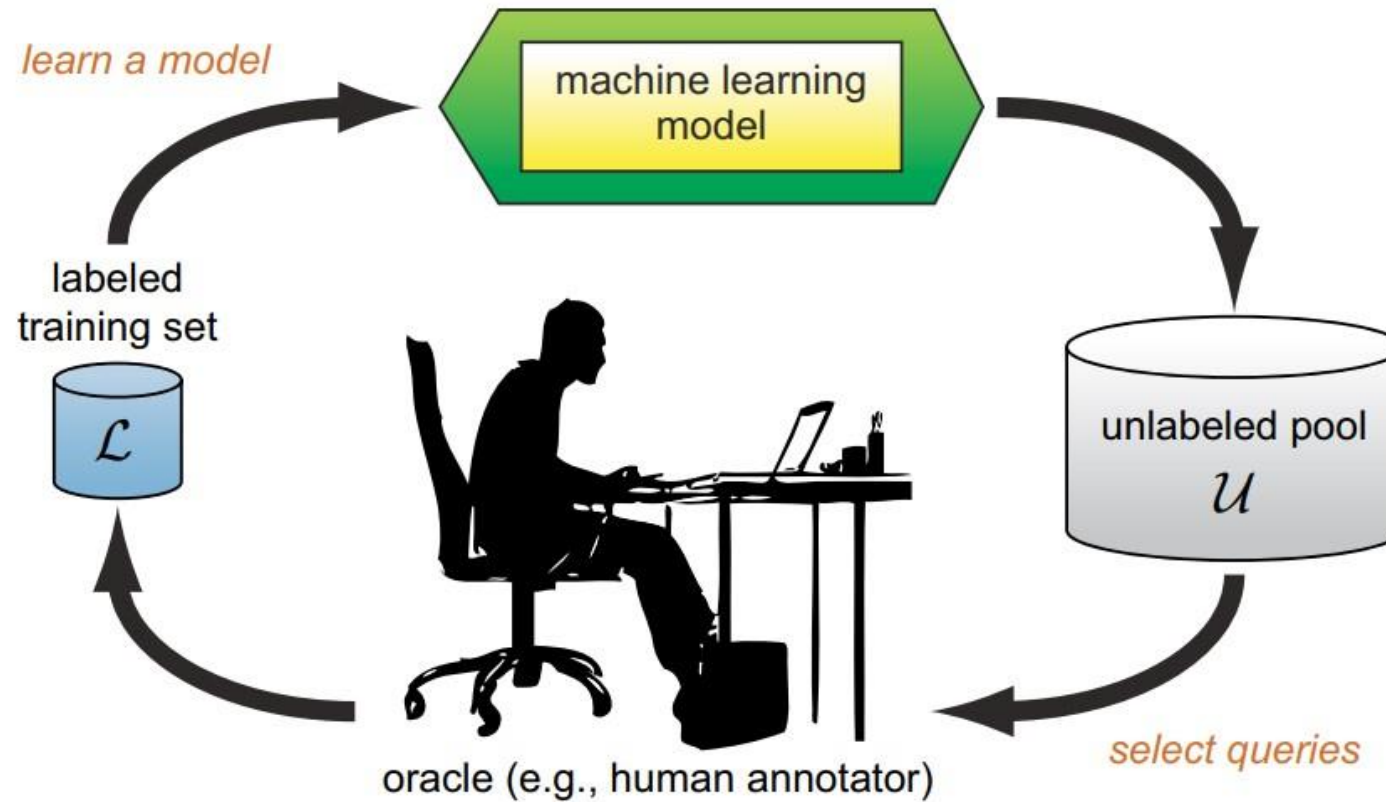


*Too expensive!*

**Active Learning:**  
Estimate which points are most valuable to solicit labels for

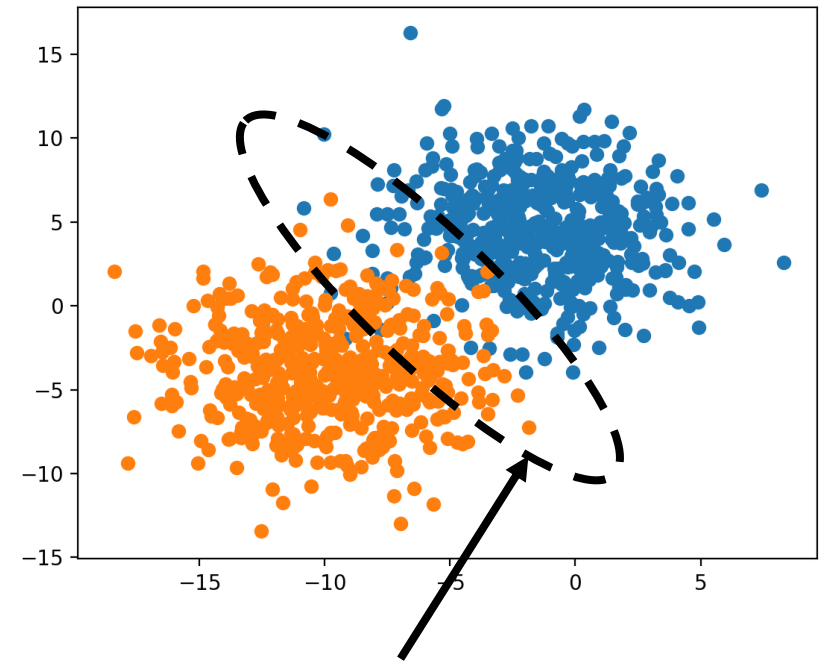
We'll discuss this today

# Active Learning: Human-in-the-Loop



# Active Learning – An Example

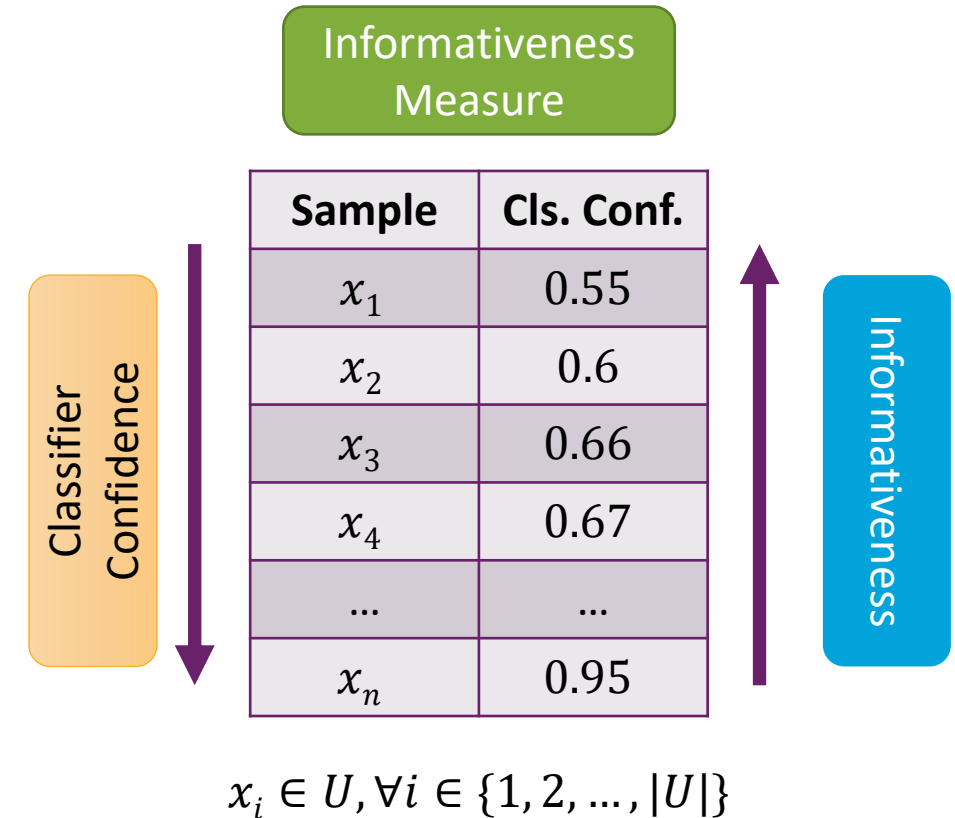
- Assumption
  - We've huge sample of *Unlabelled data (U)*
- Aim
  - Oracle should label *I*
- **Step 0:** Divide *U* into
  - *Seed (L)*: some labelled samples from *U*
  - *Unlabelled pool (U)*: rest of the dataset



Most informative area (*I*)

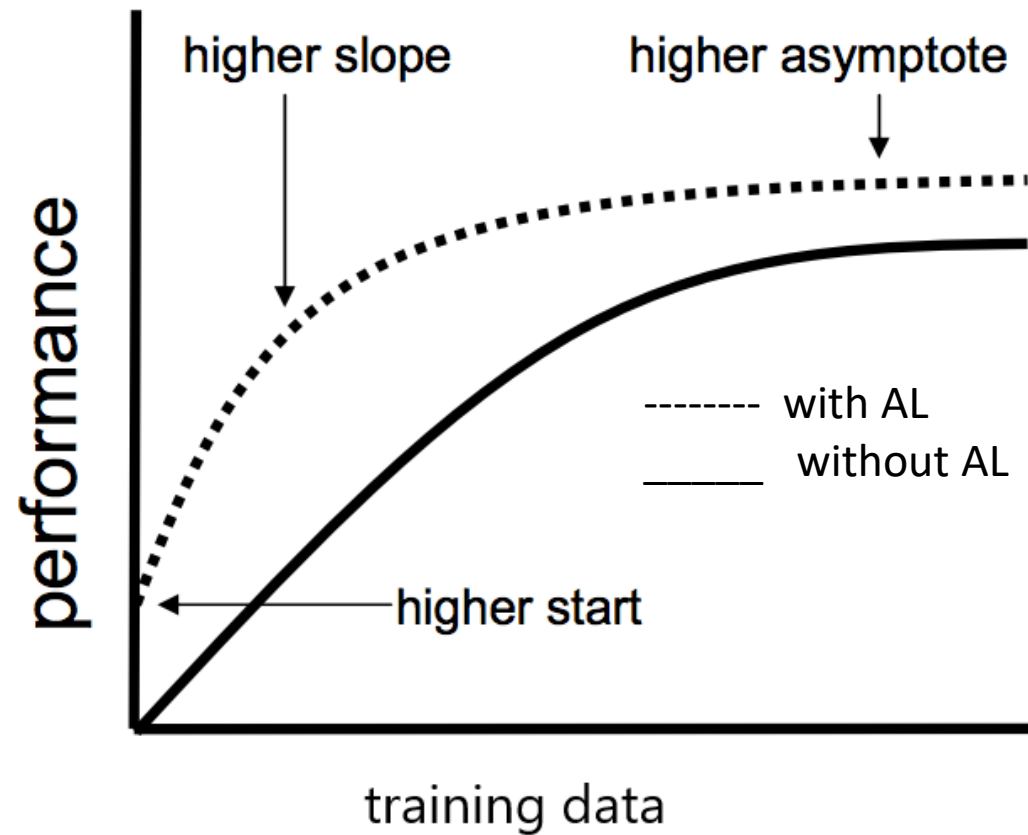
# Active Learning – An Example

- **Step 1:** Train a *classifier* ( $C$ ) on  $L$ 
  - Classifier based on domain knowledge
- **Step 2:** Use  $C$  to rank samples of  $U$ 
  - Rank based on some *informativeness*
- **Step 3:** Label *informative samples*
- **Step 4:** Add new labelled samples to  $L$
- Repeat until stopping criteria





# Active Learning – Goal



# Closing Notes

- Thumb rule of Machine Learning
  - Data first, algorithm next
  - Always try Active Learning for manual labelling
- Experience in Low Resource ML on your resume
  - Always set you aside from the crowd
- Must Read:
  - Settles, Burr. **Active learning literature survey**. University of Wisconsin-Madison Department of Computer Sciences, 2009.

# Thank You!

- Please feel free to ask any question
- Reach me at:
  - E-mail: [ashish.{rgu/iiitnr}@gmail.com](mailto:ashish.{rgu/iiitnr}@gmail.com)
  - LinkedIn: <https://www.linkedin.com/in/ashishu007/>
- Want to know about my research?
  - I blog at: <https://ashishu007.live/blog/>
  - Soon, I'll write this talk as well