



WOMEN IN DATA SCIENCE



#WiDS2020



#WiDS2020



WOMEN IN DATA SCIENCE

Smart Data and AI Life Cycle Management

智慧的数据助力AI生命周期管理

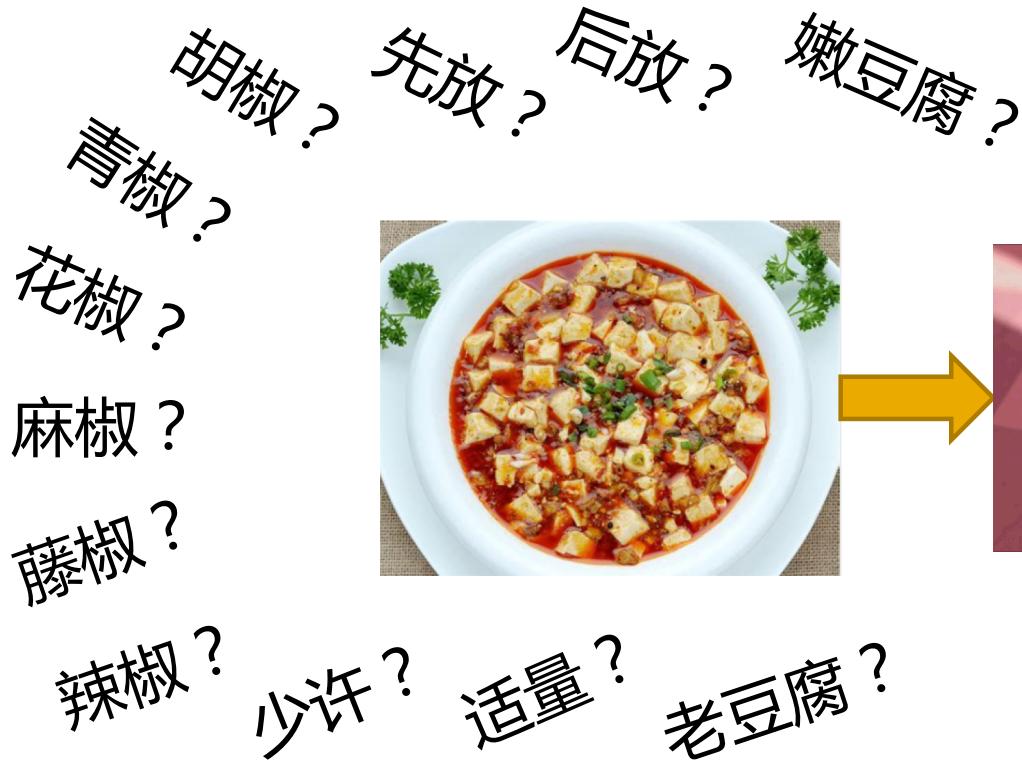


Qing WANG 王青
Research Staff Member, IBM Research China



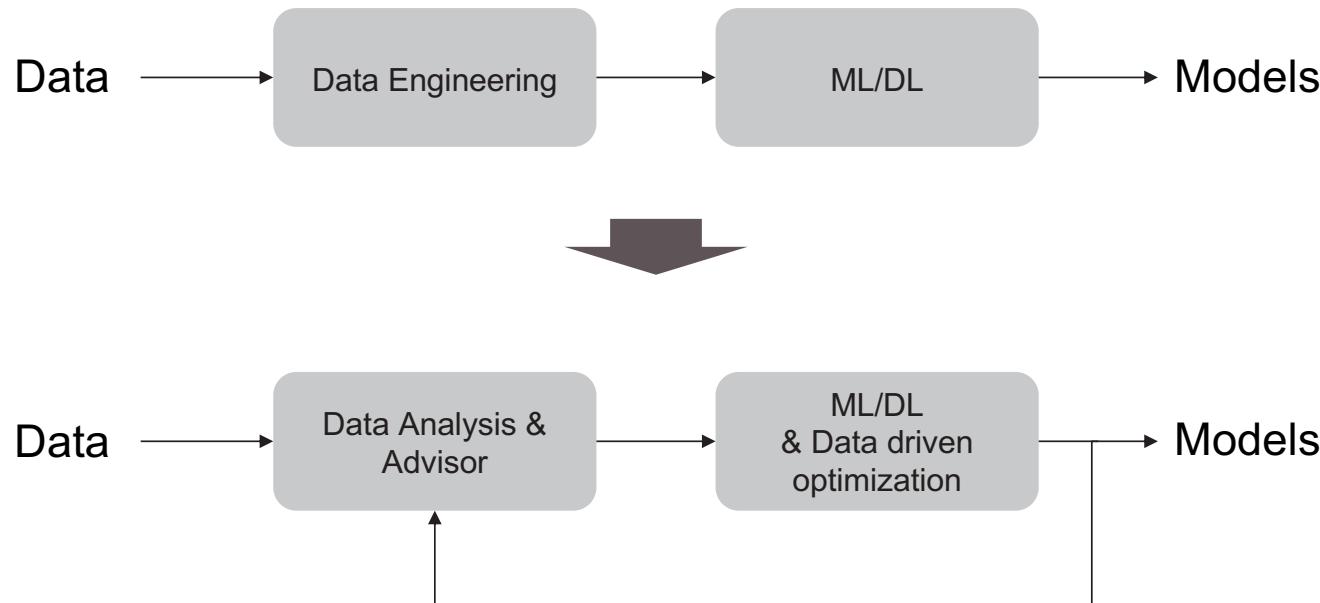
WOMEN IN DATA SCIENCE

We process data everyday.

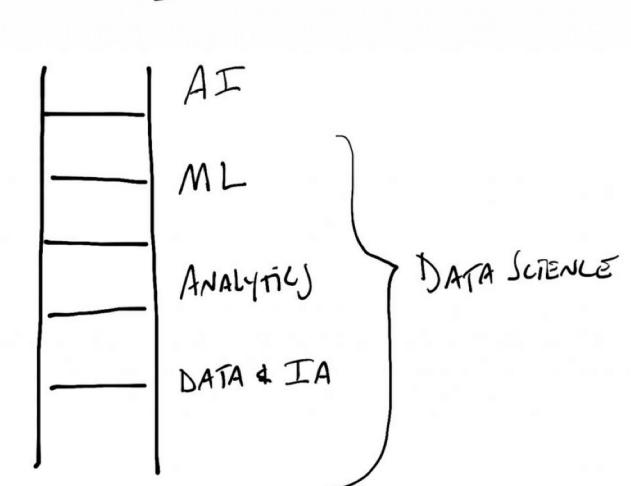


WOMEN IN DATA SCIENCE

Data for AI



AI LADDER



Ref: <https://www.ibm.com/blogs/think/2018/02/ibm-ai-ladder/>



WOMEN IN DATA SCIENCE

What's Smart Data?

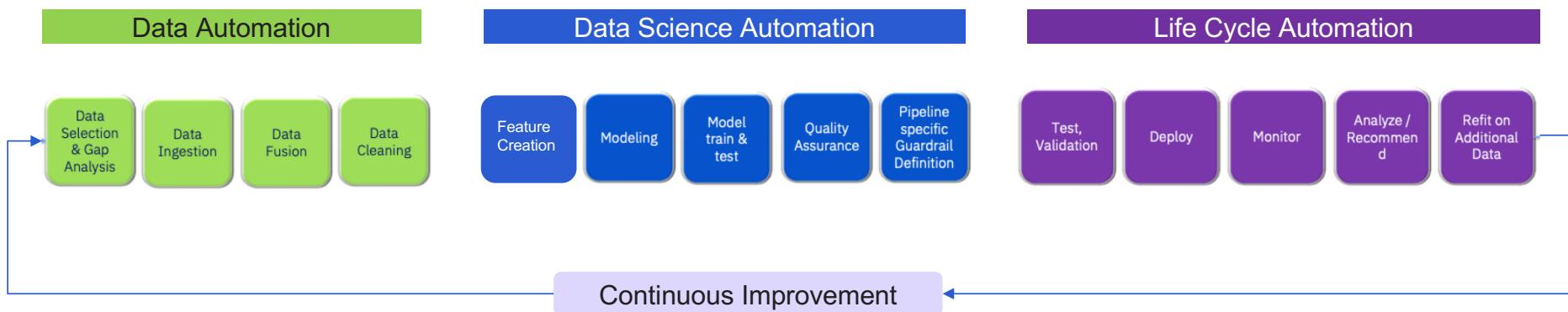
数有所用
数尽其用

- ✓ Simple
- ✓ Accessible
- ✓ Insight
- ✓ Acceleration
- ✓ Privacy



WOMEN IN DATA SCIENCE

AI Life Cycle Management



Data Governance



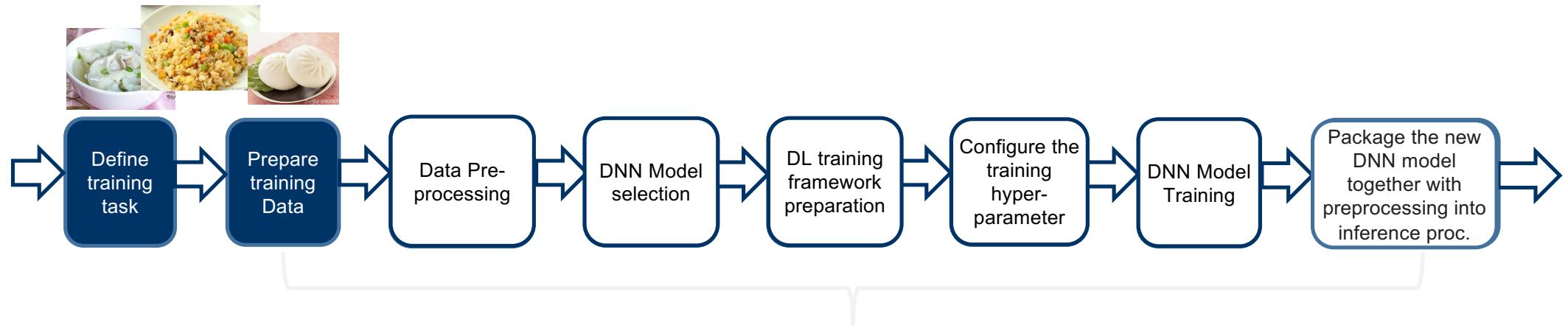
Insight Analysis & Improvement



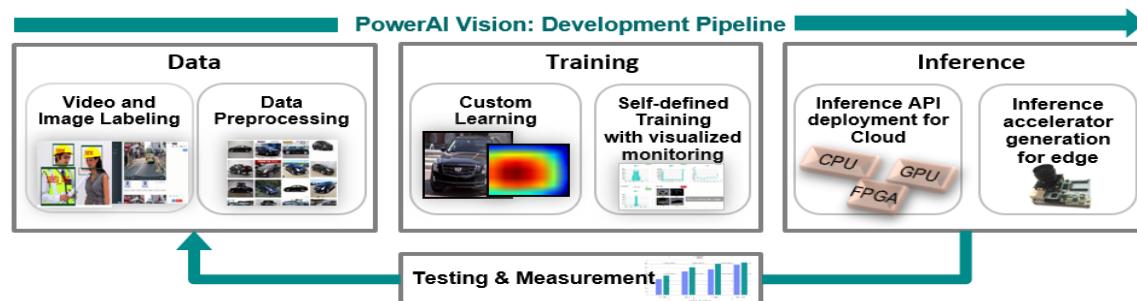
Trust and Compliance



An Example: Food Classification with Data

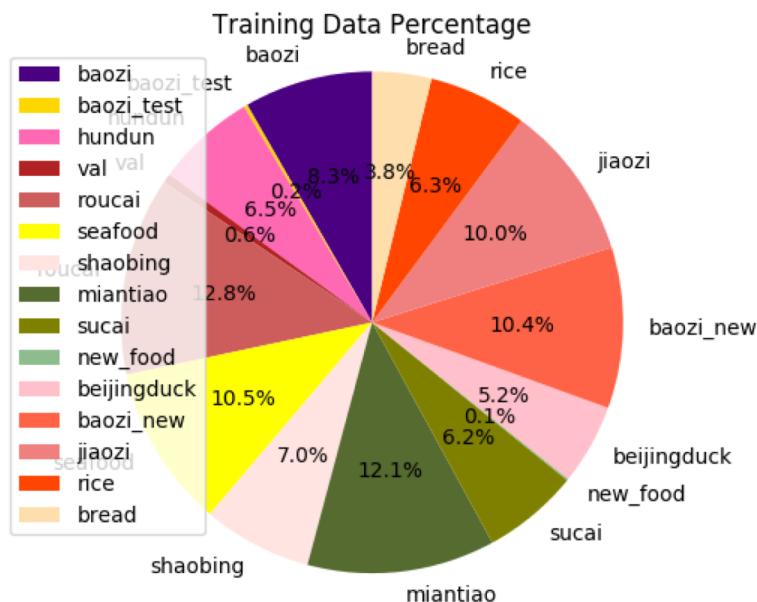


Refine and Refit Data and Models

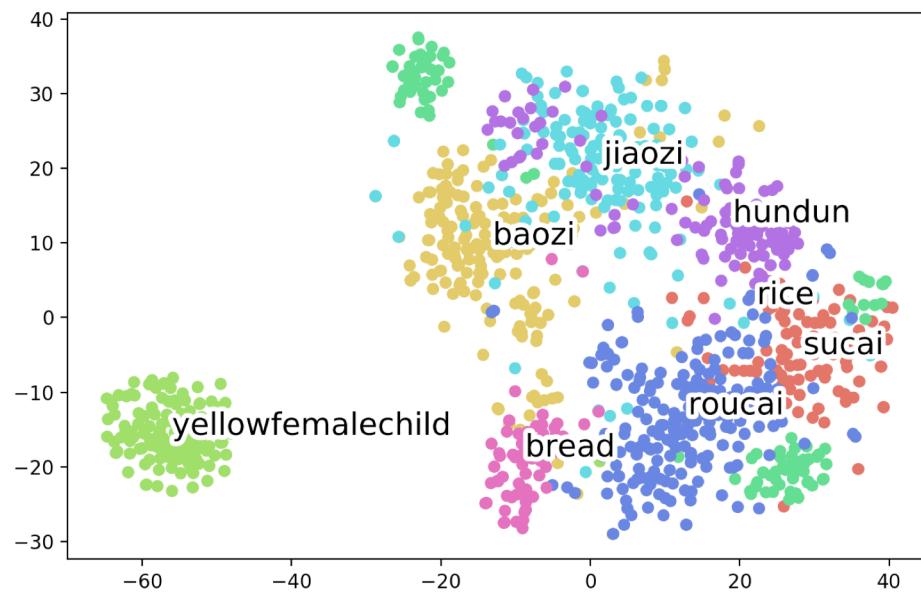


An Example: Food Classification with Data (2)

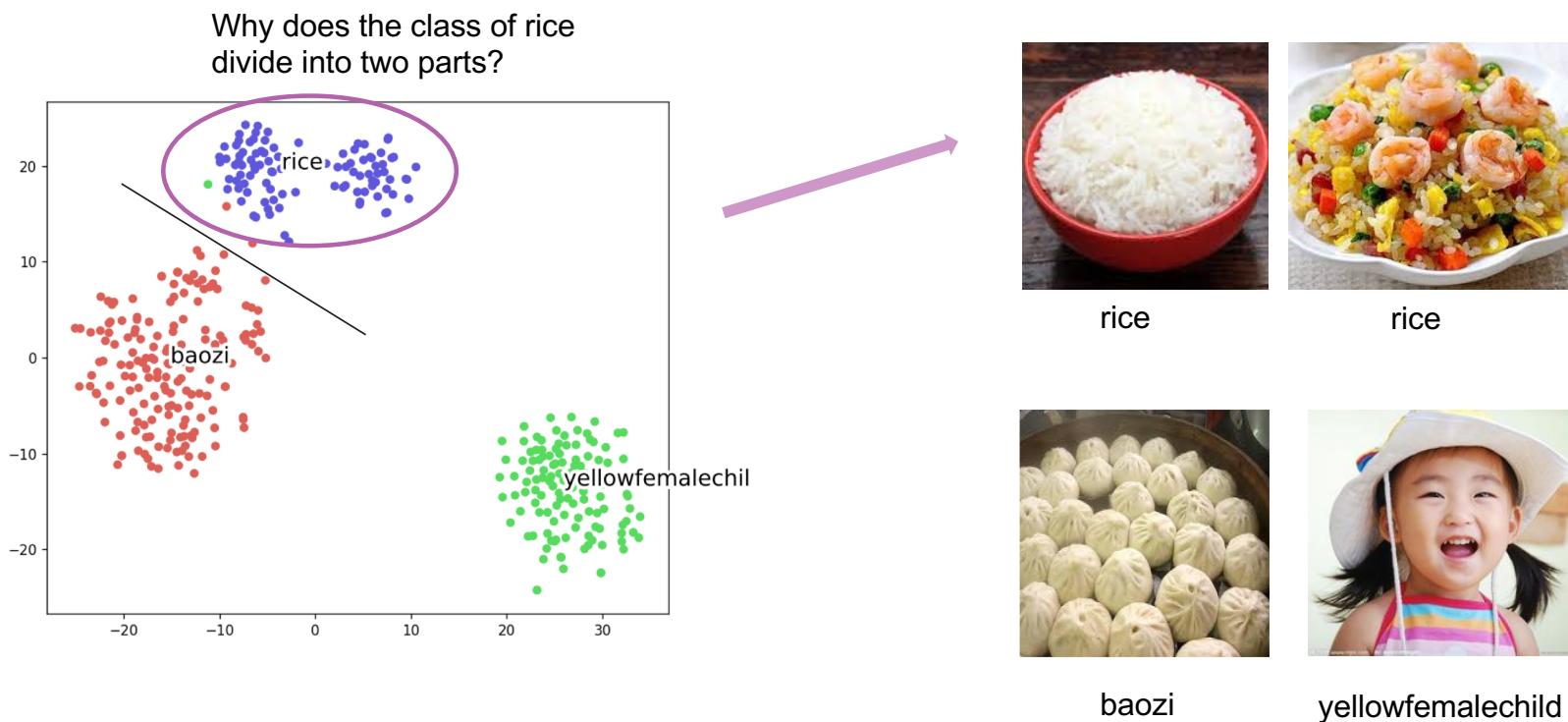
Statistical Analysis



Analysis in Feature Space



An Example: Food Classification with Data (3)



Data Automation

Data Discovery

- Data collecting,
- Data cleansing,
- Data loading,
- Data transforming
- Data mining,
- Data visualization

Data Augmentation

- Augmentation selection
- Augmentation for time/non time series data

Data Labeling

- Auto-labeling tools
- Labeling cross checking
- Error labeling removing
- Label correlation

Data Synthesis

- Cost aware
- Goal aware
- Great coverage of features and labels

Data Testing

- Evaluate data quality, such as data sparsity, bias, data imbalance
- Select data for train/val/test



Smart Data Accelerates Your AI Projects.

