Data science: competition to beat humans

CDA 数据分析师 www.cda.cn

Saeed Aghabozorgi
Data Scientist
IBM Analytics Platform, Emerging Technologies
saeed@ca.ibm.com
@SaeedAghabozorg

#BDUmeetup @bigdatau

BIG DATA UNIVERSITY

Agenda

- Human, Machines and Data Science
- Artificial intelligence
- Data Science and AI
- Data Science and AI: Challenges and opportunities
- Future


CDA 数据分析师
www.cda.cn

Human, Machines and Data Science

Human

Machines


Reliablity
CDA 数据分析师
www.cda.cn

Examples

IBM, Deep Blue

IBM Watson, Jeopardy

Google, AlphaGo

Google, self driving cars

IBM's Watson, Cancer

BIDU, Standford, Speech

Artificial intelligence

- Traditional Paradigm: Rules, methods, …

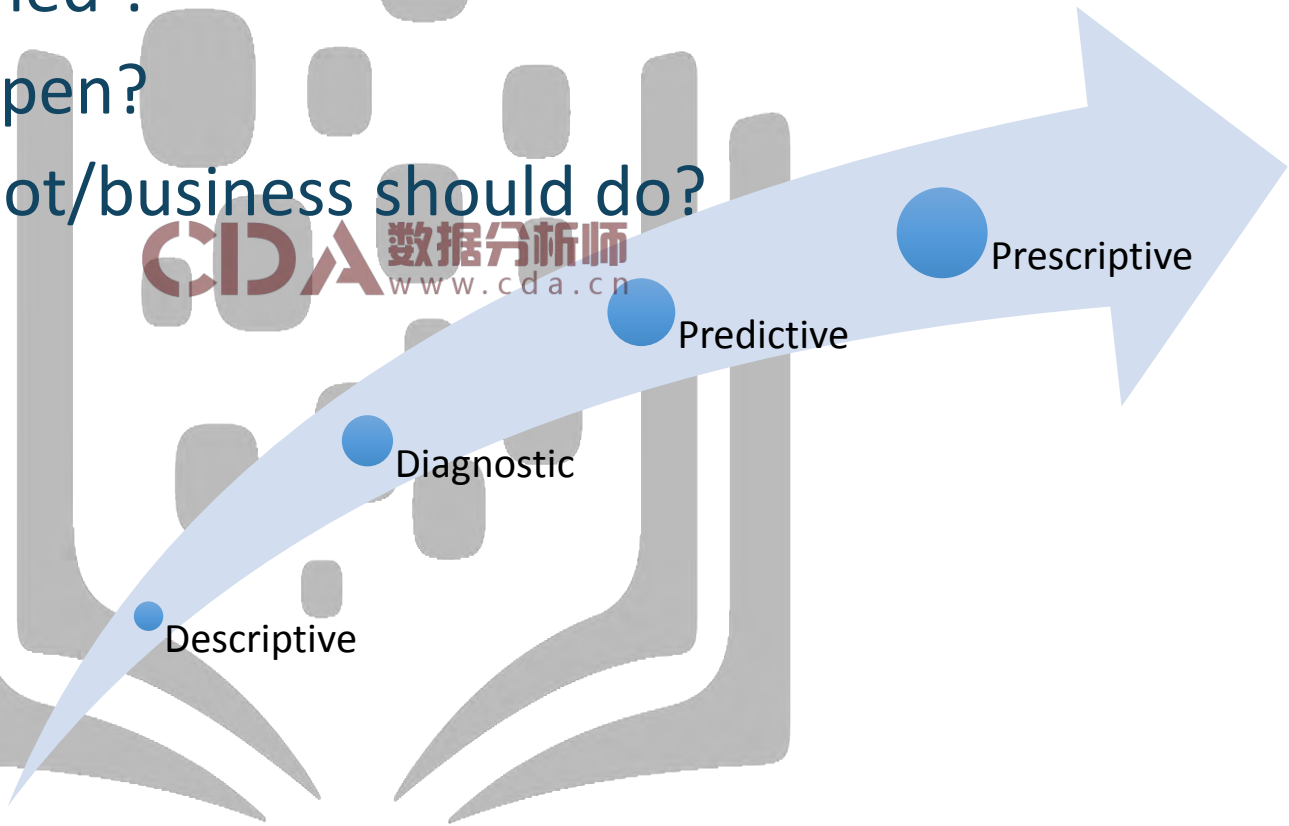
- AI components:
  - Computer Vision
  - Language Processing
  - Creativity
  - Summarization

- Revolution in AI:
  - Machine learning

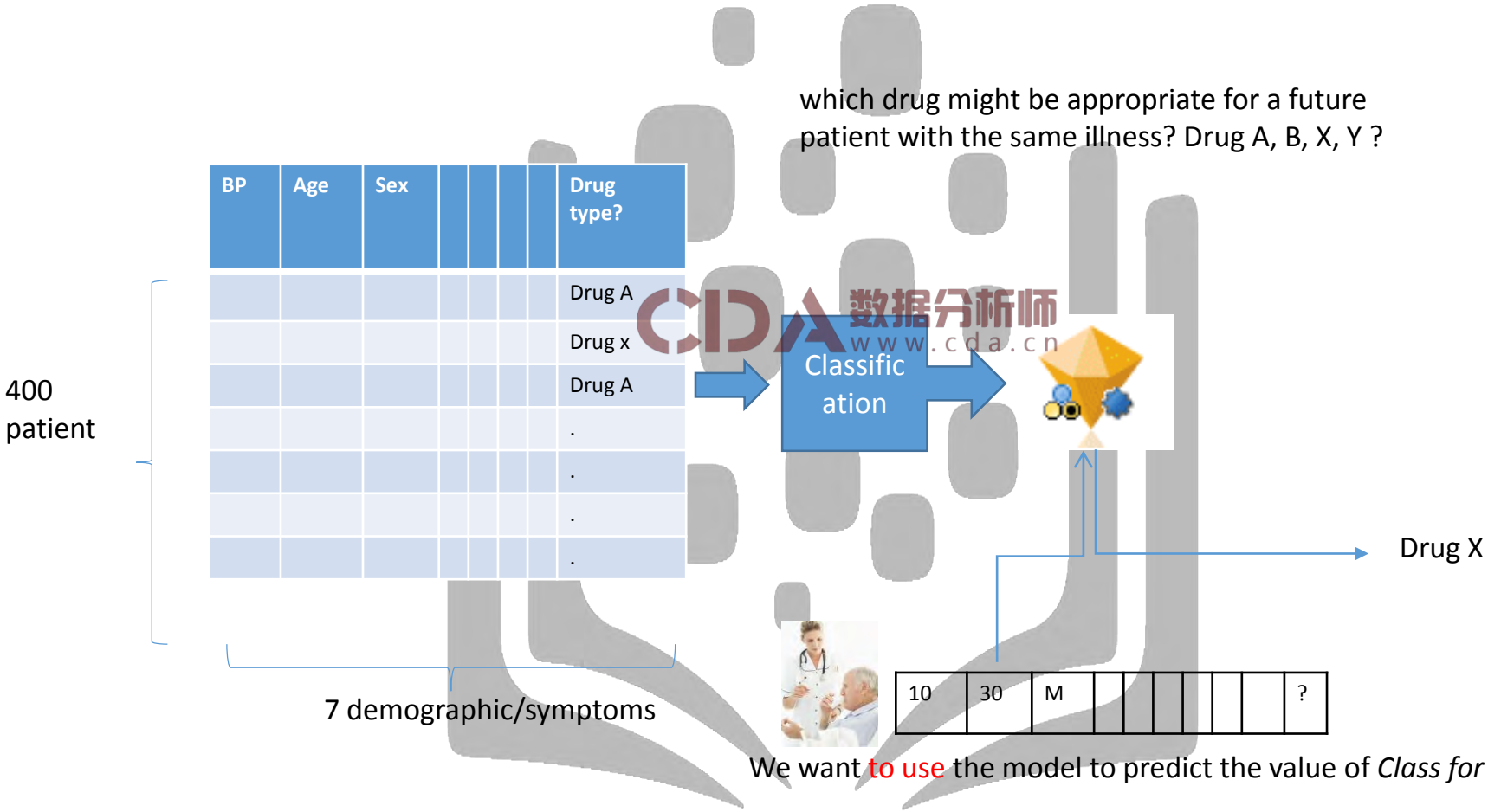
- Augmented AI:
  - by data science

So, how data science can AI?

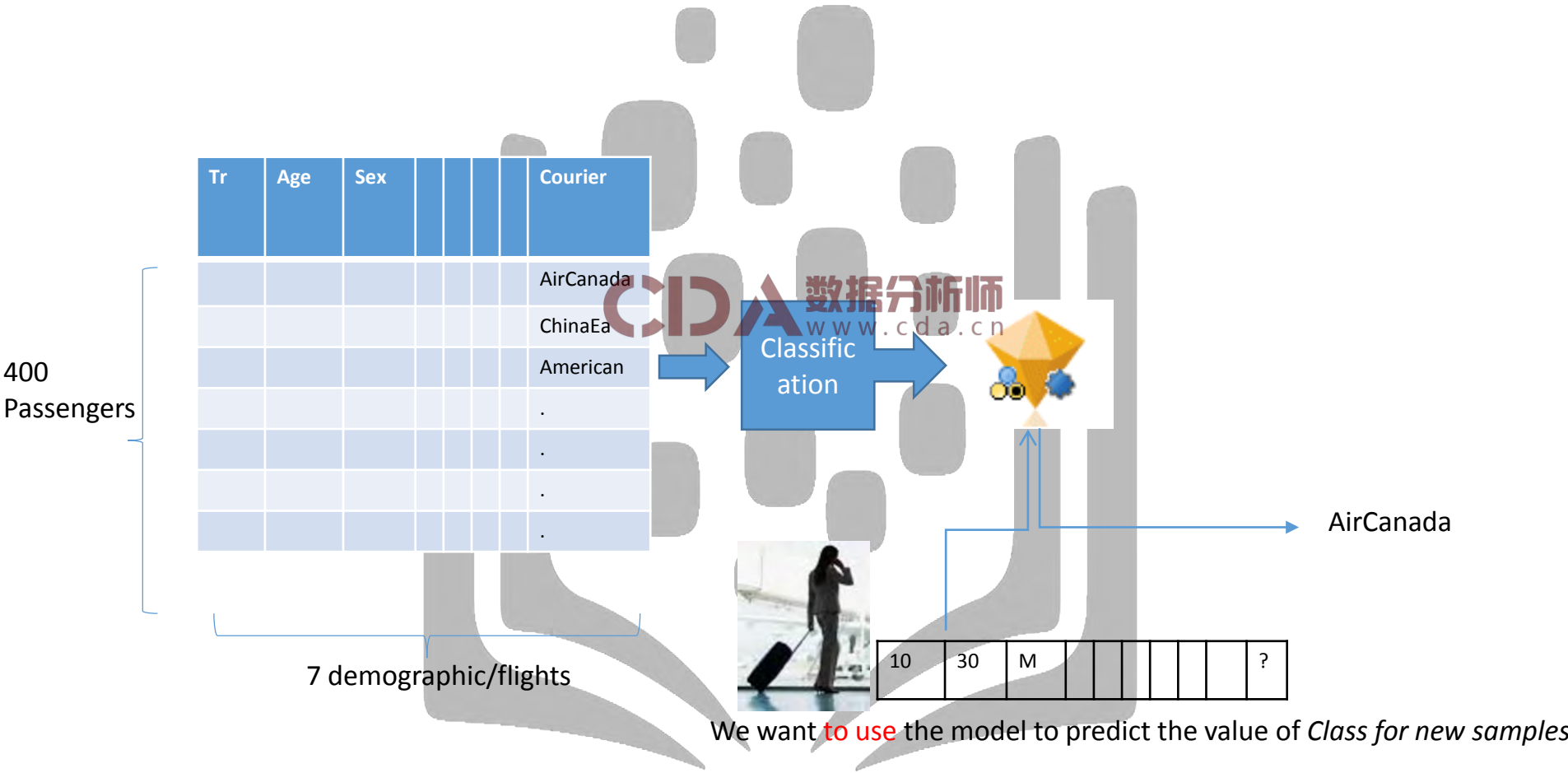
- What has happened in past?
- Why it happened ?
- What will happen?
- What we/ robot/business should do?



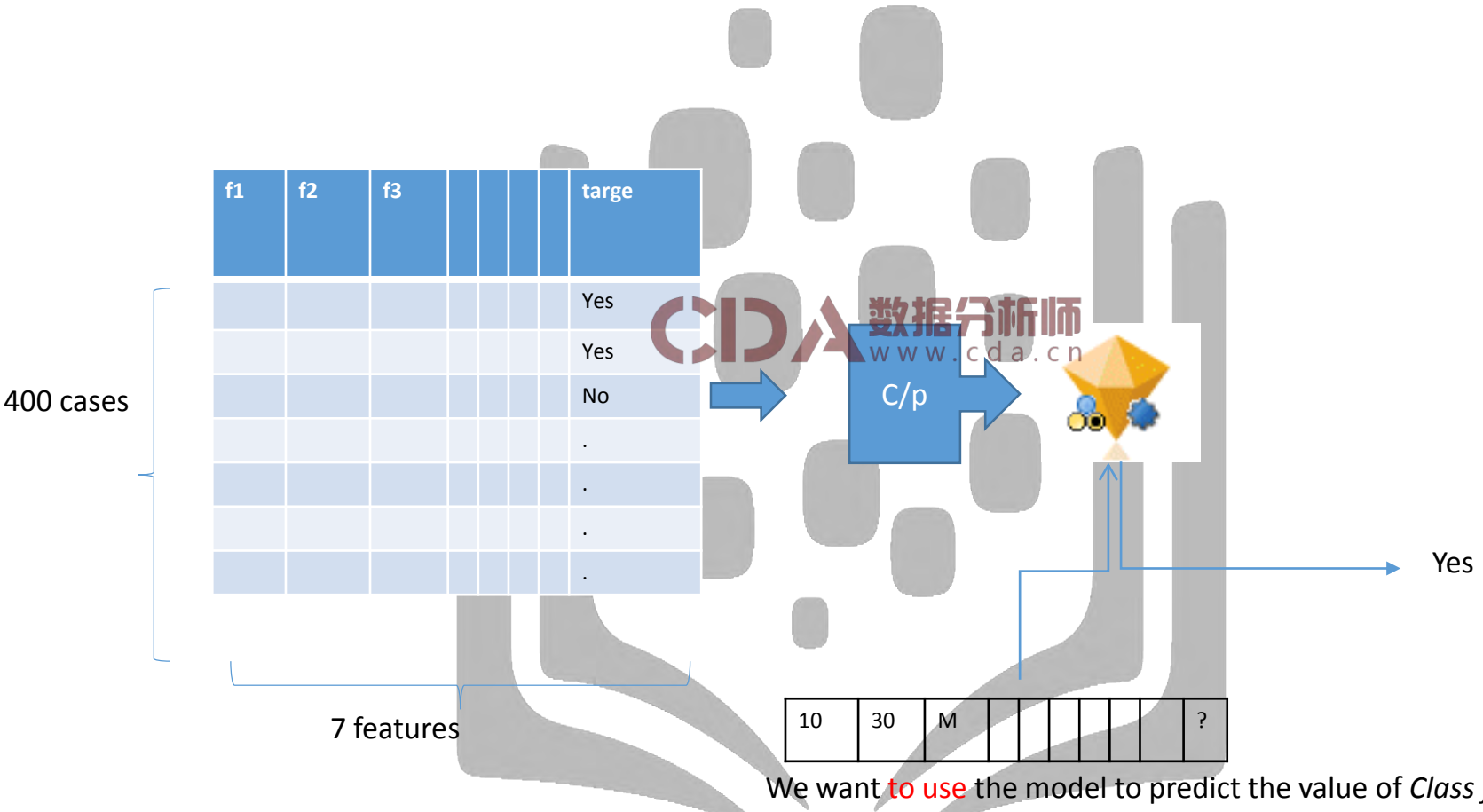
How data science can help AI?



How data science can help AI?



how data science can help AI?



Challenges in AI

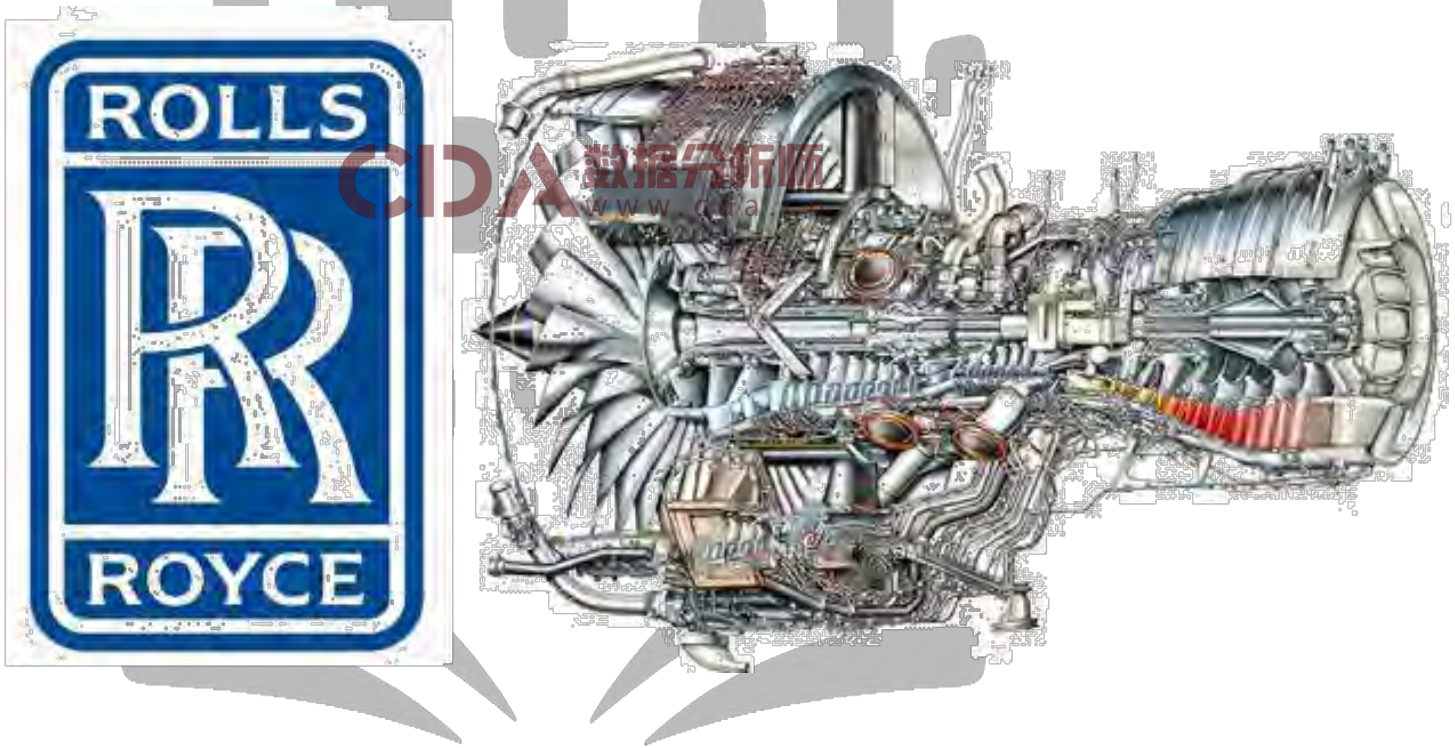
- Manual data collection
- Manual data pre-processing
- Unstructured data
- Feature engineering or use of domain specific knowledge
- What kind of data can be added to make it better?

Challenges in AI

- Which **techniques** to use for modeling?
  - Experience
- Which **algorithm/s**?
  - Insurance
- **Solutions vary from one case to another**
  - Every medical case is different.
- **Solutions change over time.**
- **How to combine different models?**

Challenges in AI

- Real-time analysis
  - Strong pre-processing

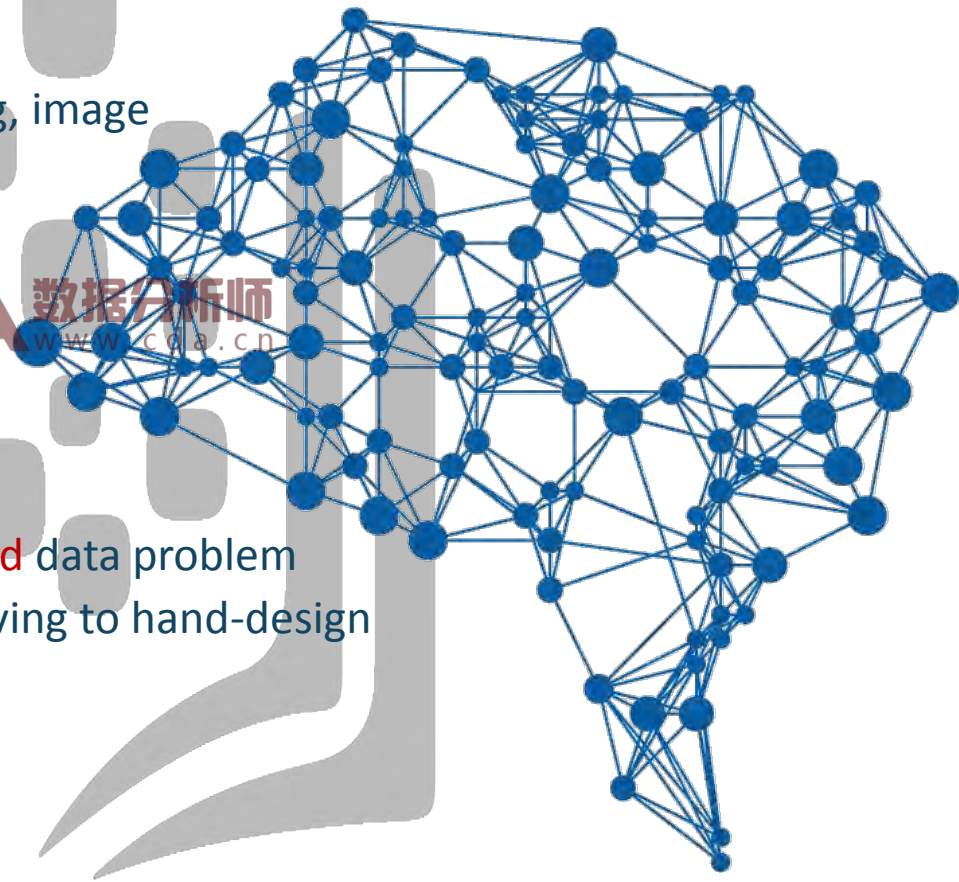


Any opportunities?

- Democratizing data
  - Open data by governments
- Advances in DS and ML, e.g. in Deep learning
  - Spark, TensorFlow
  - Deep learning
- Hardware
  - GPU, TPU,NPU, Nvidia's deep learning chip, IBM's TrueNorth
- Easy to start
  - New tools, Libraries
- Talent
  - AI startups

Techniques

- Deep Learning (what is good for?)
  - Image recognition, image captioning, image colorization
  - Text recognition, generation
  - Speech recognition, translation
  - Sound prediction



- Deep learning is important in AI:
  - To solve the variety and unstructured data problem
  - To free computer scientists from having to hand-design algorithms

Machine Learning

10	30	A	B				
20	10	X	G				
5	30	F	I				

siz	car	dri					price
10	30	A					1.5M
							3 M
							2 M
							.
20	10	X					.
							.
5	30	F					.

Deep learning

f1	f2	f3					target
a1	a2	a3					1.5M
							3 M
							2 M
							.
b1	b2	b3					.
							.
c1	**c2**	**c3**					.

Deep learning

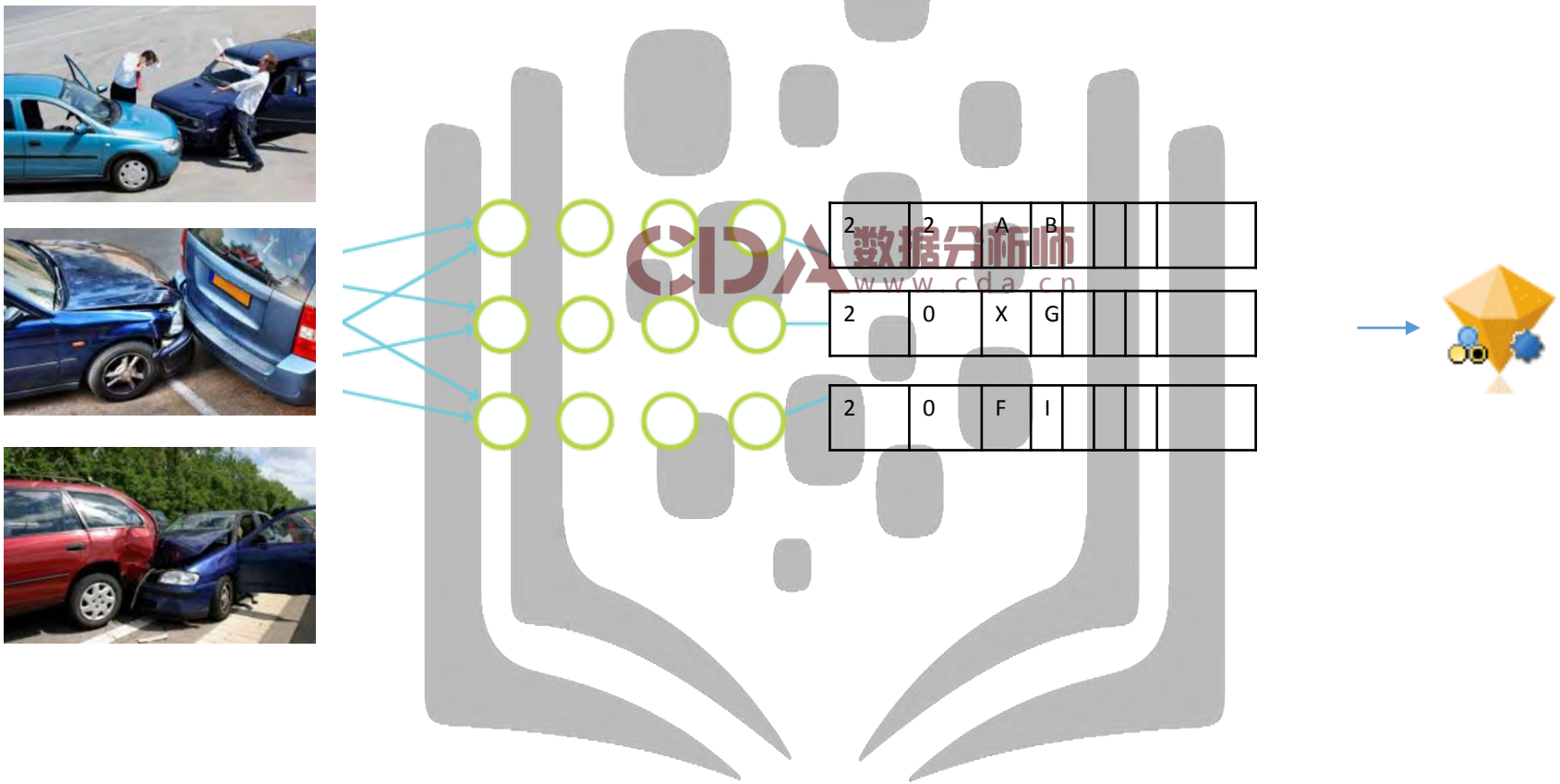


Image Classification

Classifier	Confidence Score
People	72%
Crowd of People	66%
Human	61%
Crowd	60%
Hands	60%

72	66	F	I				

IBM Watson Developer Cloud

Future of AI

f1	f2	f3					target
a1	a2	a3					1.5M
							3 M
							2 M
							.
b1	b2	b3					.
							.
c1	**c2**	**c3**					.

Future of AI

Machines

Machines

Data Scientist Workbench

www.DataScientistWorkbench.cn/

Making Open Data Science Easy

BIG DATA UNIVERSITY

www.DataScientistWorkbench.cn

BIG DATA UNIVERSITY

BIG DATA UNIVERSITY

https://my.datascientistworkbench.com

Data Scientist Workbench

發現資料
- IBM Analytics Exchange
- Open Data

管理資料
- 我的資料

準備資料
- OpenRefine

建置分析
- Jupyter Notebook
- Zeppelin Notebook
- RStudio IDE
- Seahorse

資源
- 提交您的建議
- Blog
- 知識庫
- 用戶論壇
- 技術支援

polong
登出

今天你 **想用** 哪個工具呢？

管理資料

我的資料

準備資料

OpenRefine

建置分析

Jupyter Notebook | Zeppelin Notebook | RStudio IDE | Seahorse

技術支援

了解更多有關 Workbench 的使用

OpenRefine

- 在 Big Data University 的 OpenRefine 簡介

Jupyter Notebook

- Tutorials
- 如何在 Python Jupyter notebook 中使用 SQL 處理 Hadoop 資料

RStudio IDE

- 以 R Studio IDE 在 Hadoop 叢集上執行 R 任務