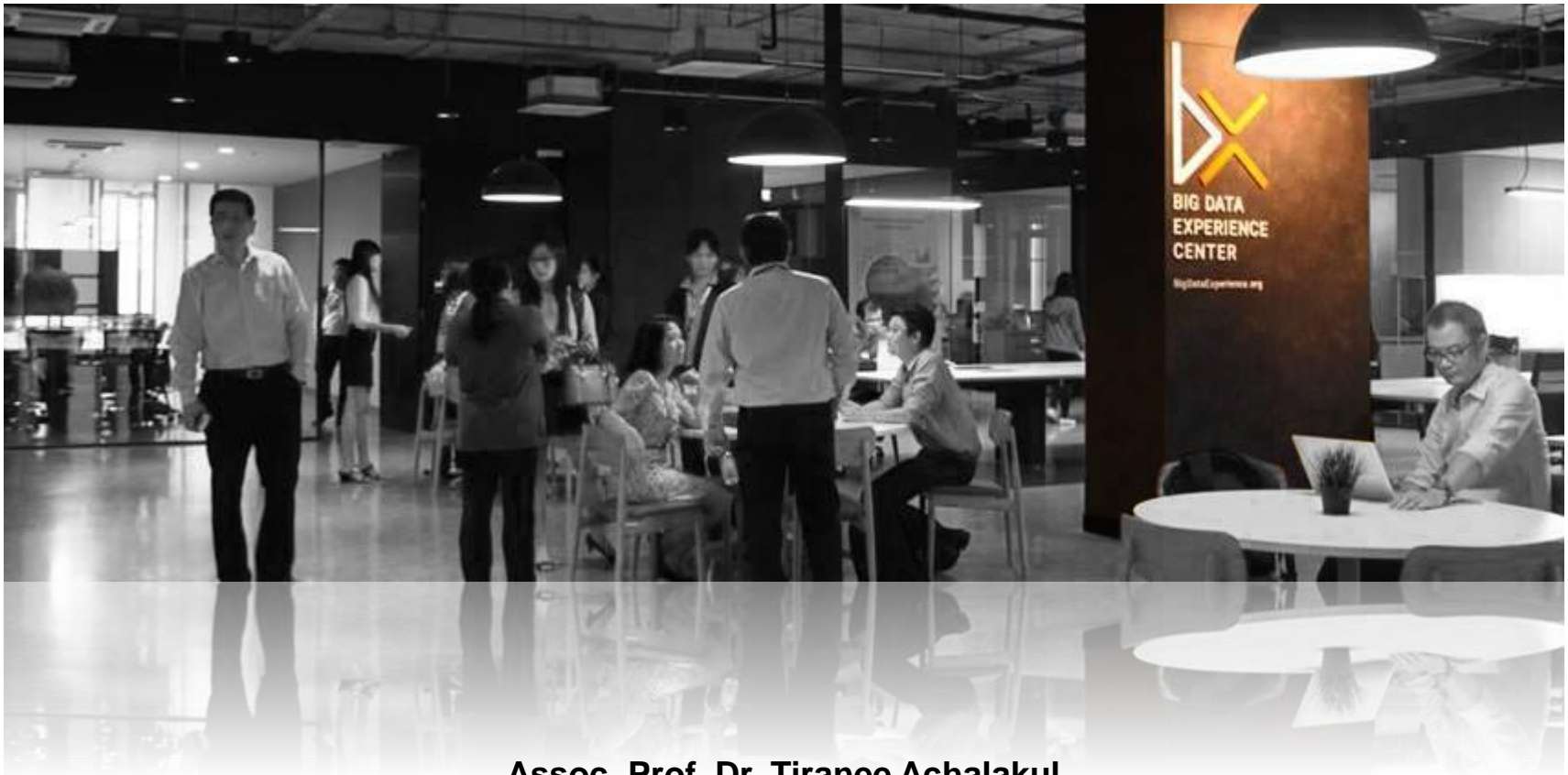




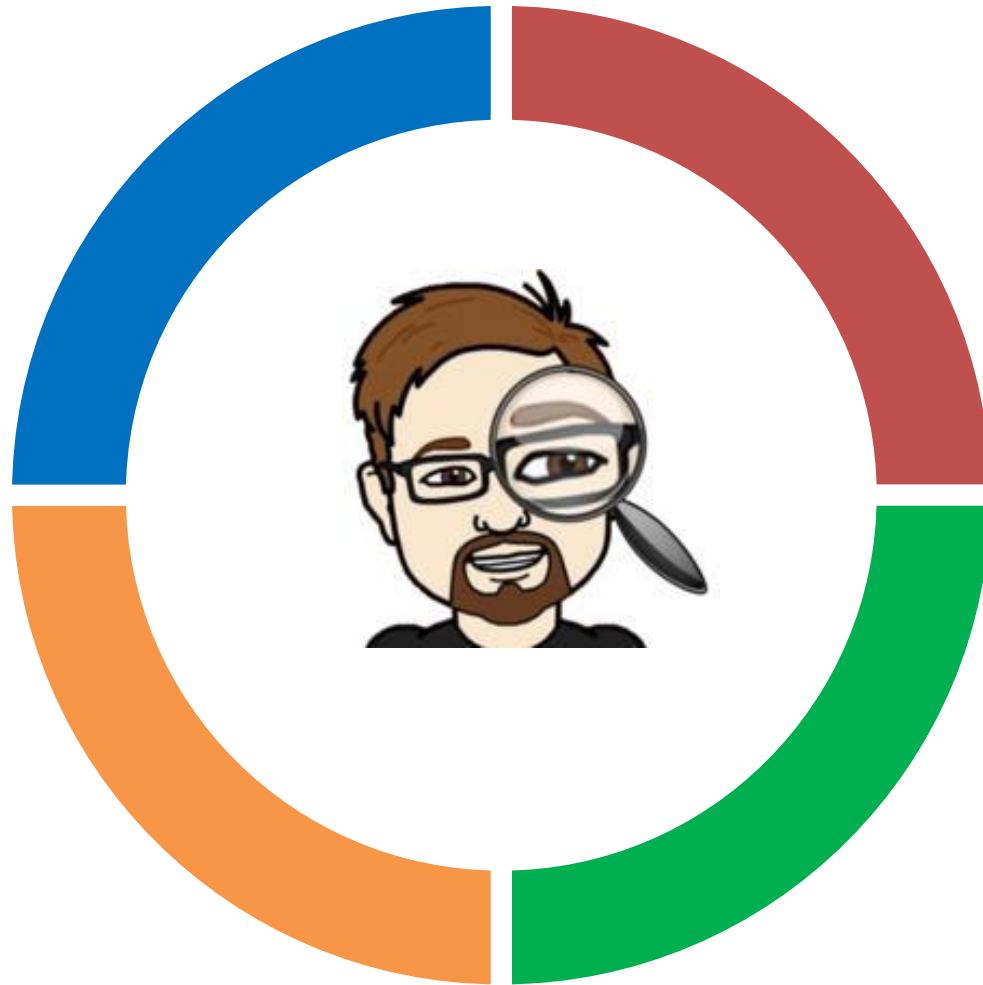
A market place where everyone is exchanging idea and practice around Big data technology; to create a practical realization of Big data in Thailand



Assoc. Prof. Dr. Tiranee Achalakul

Academic Director

BIG DATA



BIG DATA

VOLUME

Online & Offline
Automatically generated
Manually created

VARIETY

Unstructured
Structured

- ✓ **Behavioral data** - server logs, clickstream, ATM log
- ✓ **Images & sounds** - photographs, videos, Google street views images, medical images, handwriting images, voice recordings
- ✓ **Languages** - text messages, tweets, web content
- ✓ **Records** - medical records, large automated survey, tax
- ✓ **Sensors** - temperature, accelerometer, geolocation

VELOCITY

Speed of Generation
Rate of Analysis

VERACITY

Untrusted
Uncleansed
Unclear

BIG DATA ANALYTICS

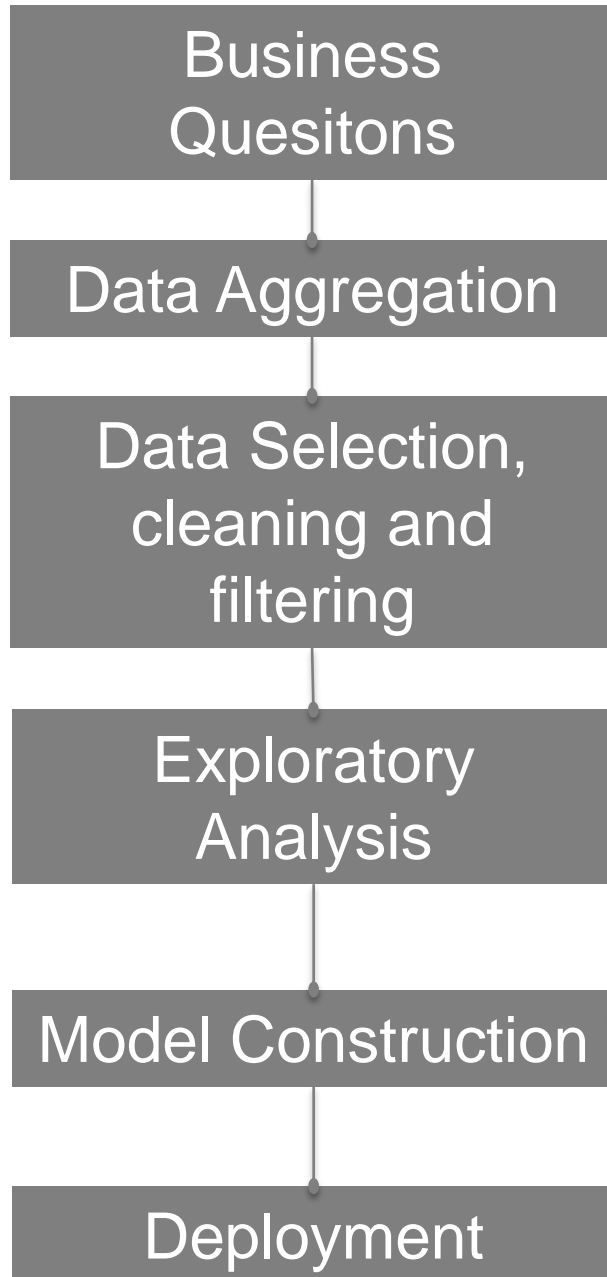
A set of fundamental concepts/principles that underlie techniques for extracting useful knowledge from large datasets containing a variety of data types. To uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful business information



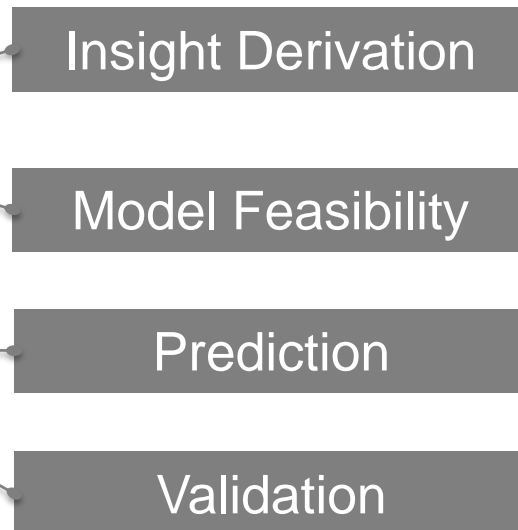


ANSWERING BUSINESS QUESTIONS

- Who are the most profitable customers?
 - A straightforward database query, if “profitable” can be defined clearly.
- Is there really a difference between the profitable customers and the average customer?
 - Statistical Hypothesis testing
- But who really are these customers? Can I characterize them?
 - Automated pattern finding
- Will some new customer be profitable ? How much revenue can I expect?
 - Predictive model of profitability



DATA SCIENCE PROCESS



METHODS



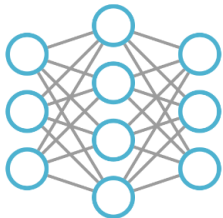
Data Mining

The Computational process of **discovering patterns in large data sets** involving methods at the intersection of statistics, machine learning, and database systems.



Text Mining

The process of **deriving high-quality information from text**. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning.



Machine Learning

The science of getting computers to learn from data without having to be explicitly programmed by humans. Machine model can teach themselves to grow and change when exposed to new data.



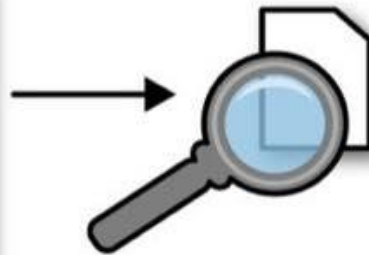
DATA MINING TECHNIQUES

- **Clustering** : Group individuals in a population by their similarity (not driven by any specific purpose).
- **Co-occurrence grouping** : Find associations between entities based on transactions involving them.
- **Profiling** : Characterize the typical behavior of an individual, group, or population.
- **Link Prediction** : Predict connections between data items (Link should exist at what strength)
- **Classification** : Predict, for each individual in a population, which of a set of classes this individual belongs to.
- **Regression** : Produce a model that, given an individual, estimates the value of the particular variable specific to that individual.
- **Similarity matching** : Identify similar individuals based on data know about them. Similarity underlie solutions to other tasks.

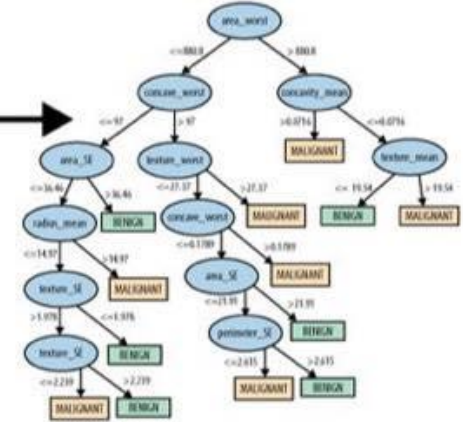
Historical Data

x	y	z	class
14	True	Red	accepted
6	True	Blue	rejected
...			
50.3	False	Red	accepted

Data mining



Model



Training data have all values specified

Model is deployed

Mining

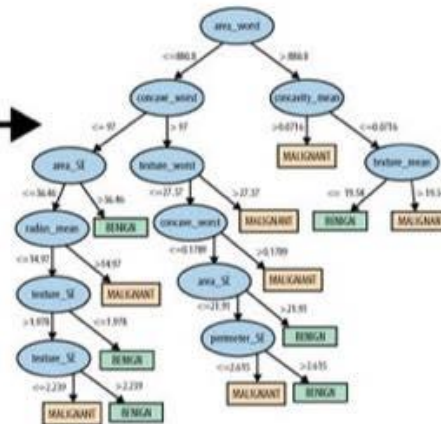
Use

New data item

x	y	z	class
30	false	Red	?

New data item has class value unknown (e.g. will customer accept?)

Model

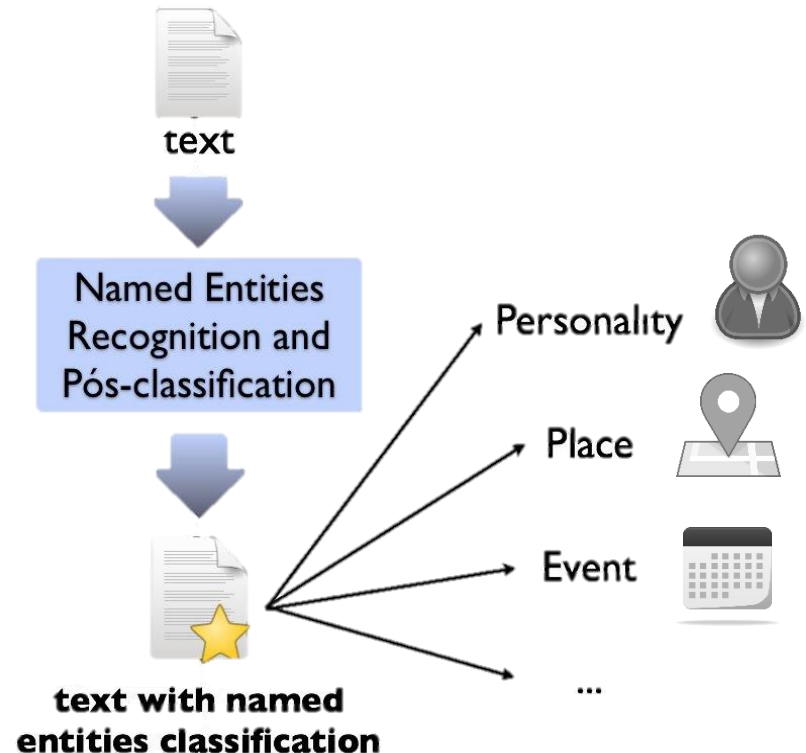


Class: accepted, Probability: 0.88

NAMED ENTITY RECOGNITION (NER)

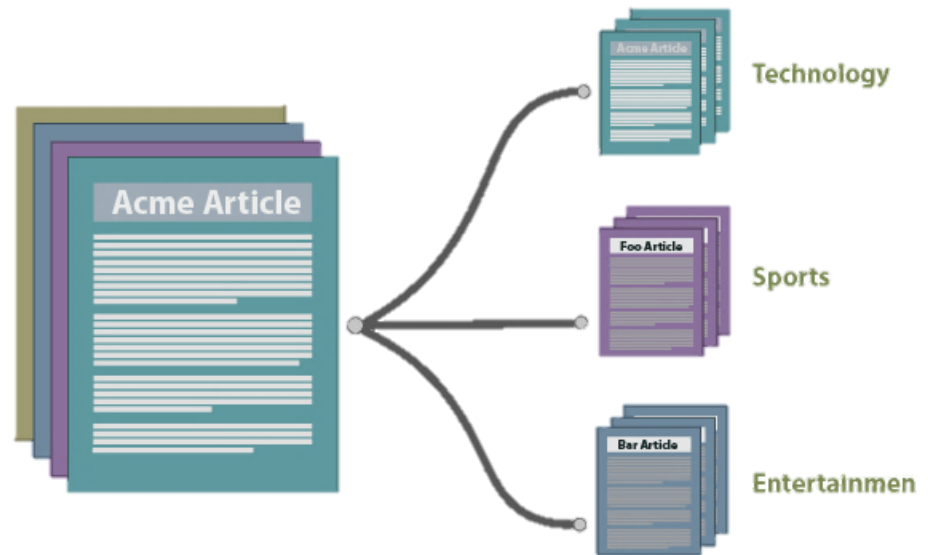
To locate and classify named entities in text into pre-defined categories such as:

- Names of persons
- Organizations
- Locations
- Expressions of times
- etc.



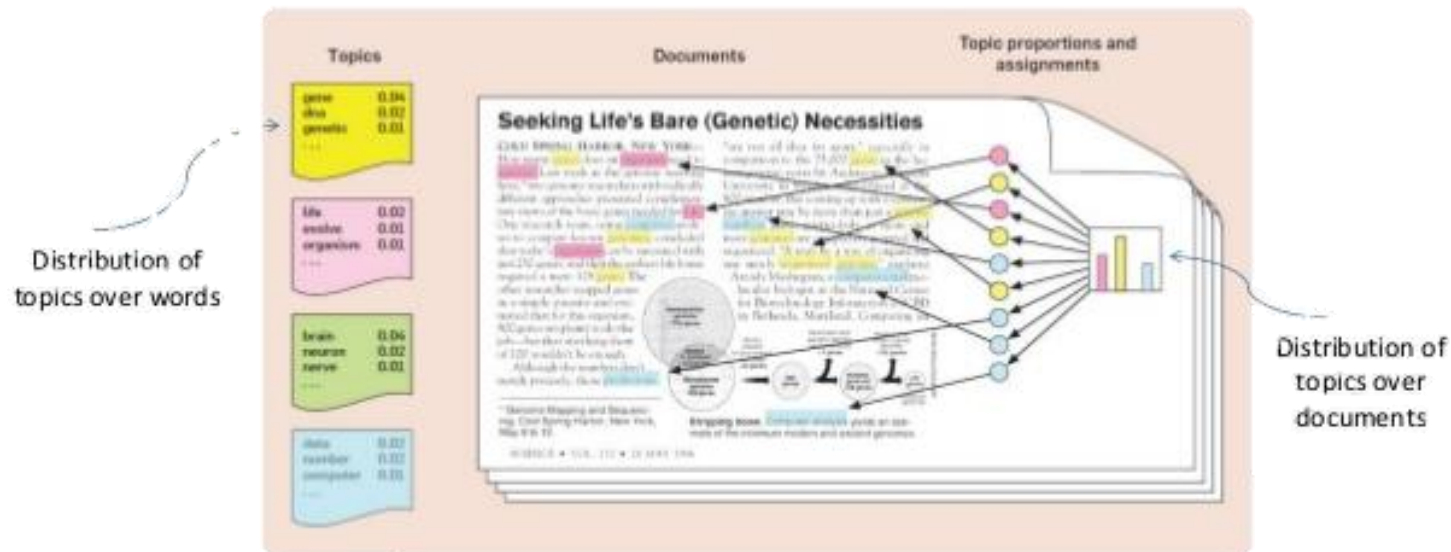
TEXT CLASSIFICATION & CLUSTERING

- Classification
 - ✓ To assign a document to one or more classes or categories.
- Clustering:
 - ✓ The application of cluster analysis to textual documents

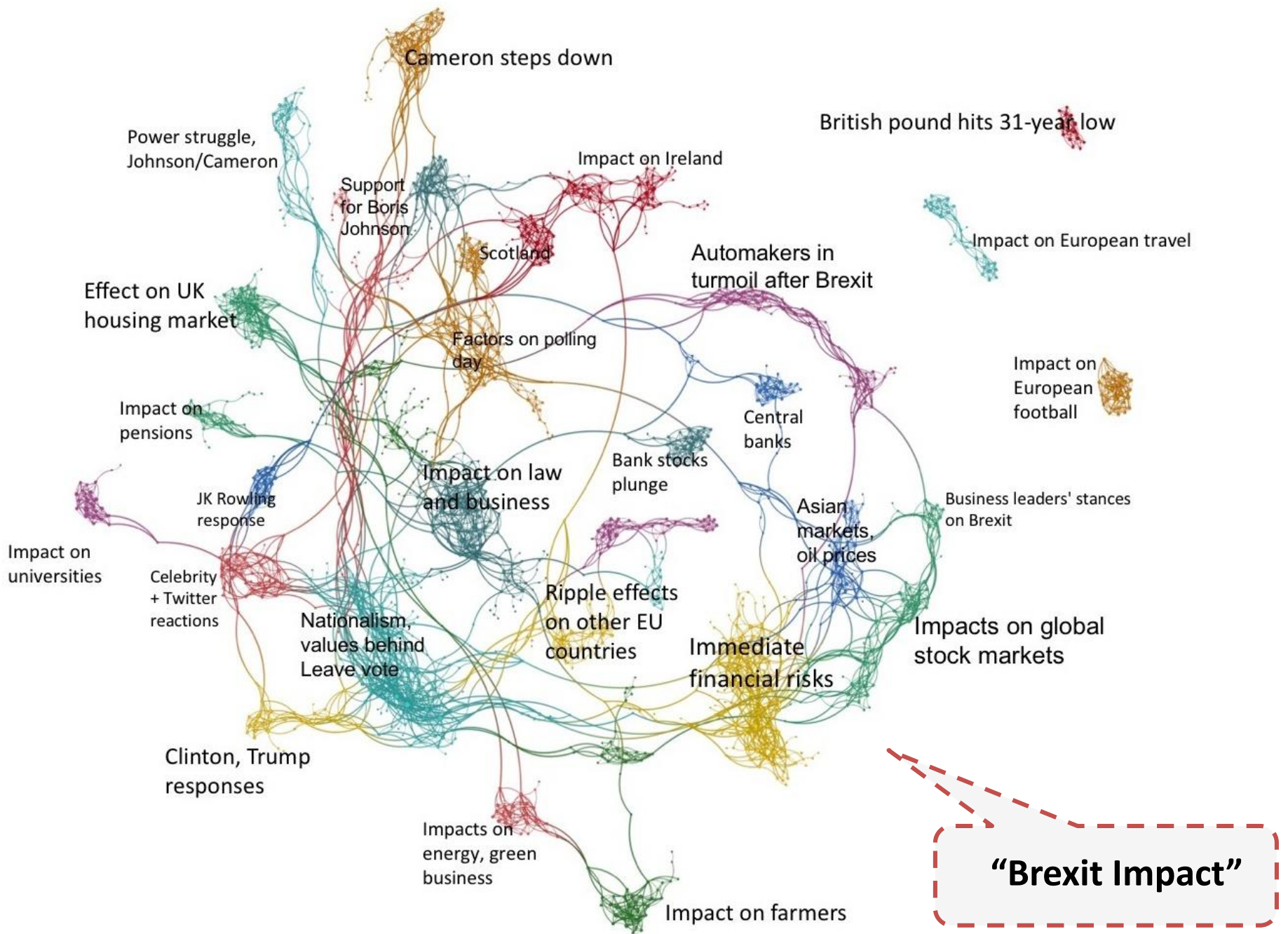


TOPIC DISCOVERY

- Characterizes document according to topics
 - ✓ Discover topics mentioned about “ประชาชาติ” on the social network
 - ✓ Discover topics mentioned about “พฤษภาคม” on the social network

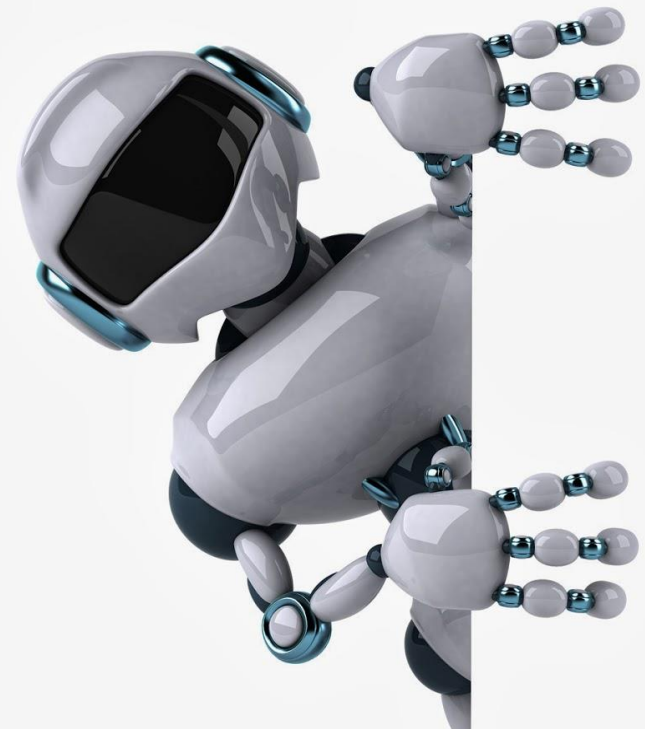


[Image from Blei, D. *Probabilistic Topic Models*, Communication of the ACM, 2012]



MACHINE LEARNING

Learns from data and make predictions about data by using statistics to develop self learning algorithm (no explicit programming)



BIG DATA AND MARKETING



Targeted market



Campaign combinations
for effective up-selling



Recommendations
for cross-selling



Predicting
customer
churn



Customer behavior analysis: the key to marketing in this digital era is contacting customers just

- When they wish to be reached
- When they are in the right location
- Then, engaging them with personalized real-time offers.



BIG DATA FOR HR

- Talent acquisition, retention, placement, promotion, compensation, or workforce and succession planning.
- Analyzing the skills and attributes of high performers in the present; build a template for future quality hiring.
- Non-traditional data gathering sources
 - Social media channels where prospective candidates usually leave their digital *'thought prints'*.
- Statistical analysis of productivity and turnover
 - Old indicators (such as GPA and education) were far less critical to performance and retention.

Bersin by Deloitte

Talent Analytics Maturity Model®



Ref: Forbe



THE CHEF BOT

Predicting ingredient-pairing and creating recipes

The key technique is to organize food in buckets and map the relationships between these sets. Sets may be based on recipes, taste, or cooks' insight.

- **IBM Chef Watson** crunches through recipes, regional and cultural knowledge and statistical, molecular and food pairing theories. Chef Watson generates new recipes with unusual ingredients.
- McCormick developed **FlavorPrint** by researching about food tastes, textures, aromas, and preparation techniques. Data then are encoded into a set of algorithms that helped select recipes.
- **Kitchology** uses a cook centric dataset (Collective wisdom of chefs). The algorithms reflect on what people who make food designs learn in culinary schools coupled with a proprietary database.

SOFIA





Facebook.com / **BigDataExperience**



CONTACT

08 9789 5224, 09 8258 8261

contact@BigDataExperience.org

JOIN

[www.BigDataExperience.org / join](http://www.BigDataExperience.org/join)

MEET

[www.BigDataExperience.org / request](http://www.BigDataExperience.org/request)

