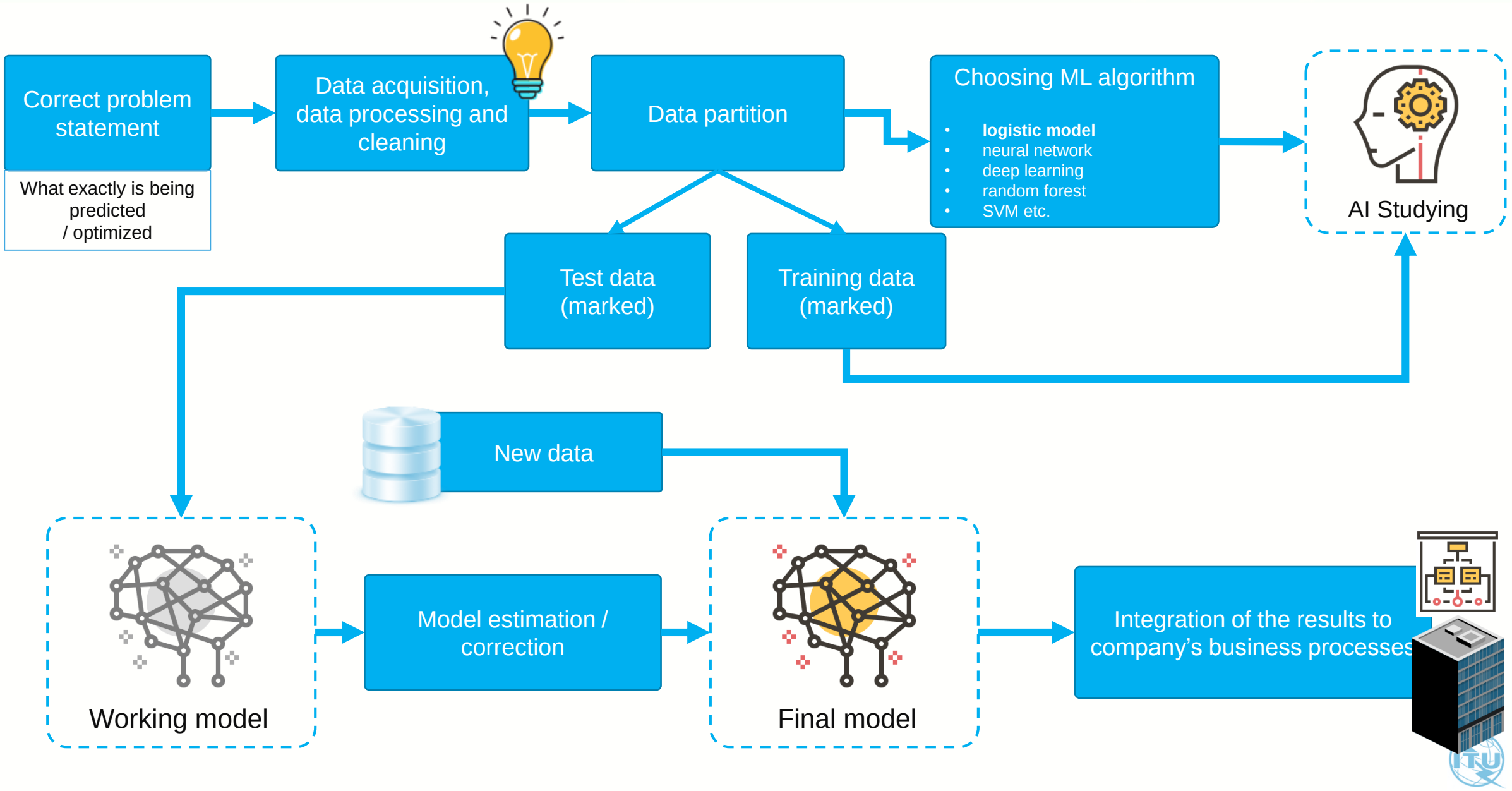

ITU

Dmitry CHASOVSKOY
Head of project department, Rostelecom (IQMEN)

Experience on Big Data processing

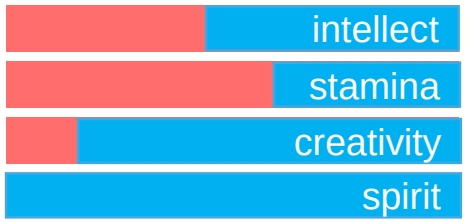
Saint Petersburg 2018

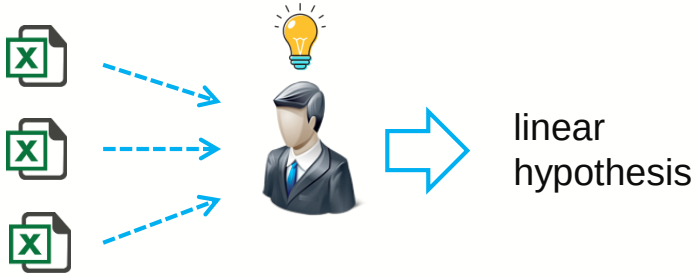
Ingredients of Machine learning



Goal: find clients in risk areas

Human - expert

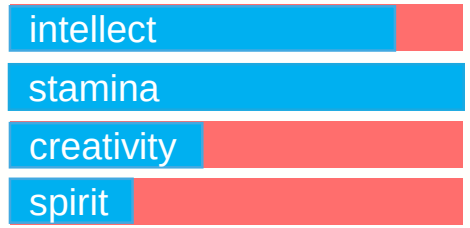




- human resources needed
- low forecast accuracy
- result depends on specific expert
- static model

VS

Artificial Intelligent



- no human resources required
- forecast accuracy is much higher
- dynamic model, can be adapted to business's changes

Forecasting the outflow of the client base

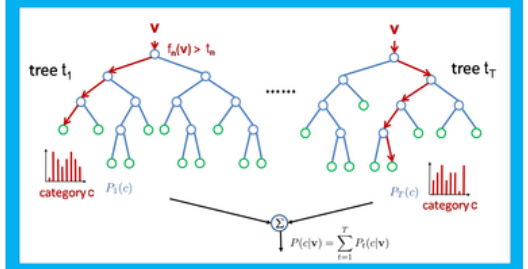
- Subscriber lifetime
- Voice and broadband traffic consumption
- Dynamics of payments
- Dynamics of account receivable
- Quantity/duration of contacts with technical support and complaints solving
- Presence of fails of broadband sessions
- etc.

Data views

- Subscribers
- Charges
- Voice/Broadband traffic
- Accounts receivable
- TechSupport contacts
- Fails of sessions

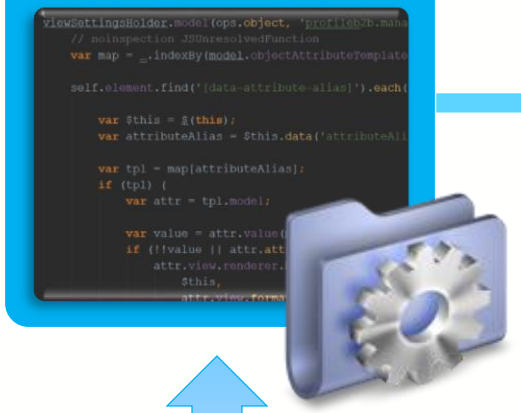
DB contains data for 2 years with subscribers outflow and their patterns of behavior

Data model Random Forest

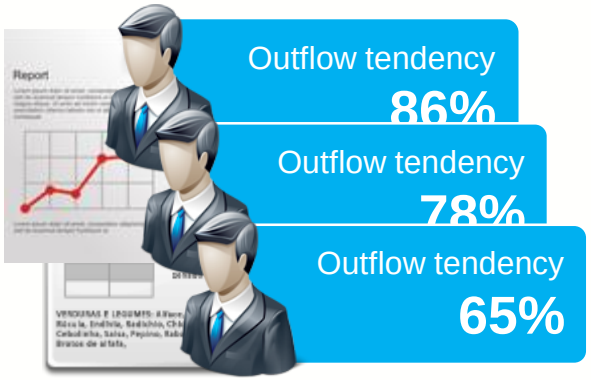


Landline, Internet, MVNO

Implementation using code



fresh data with subscribers' activity (monthly)

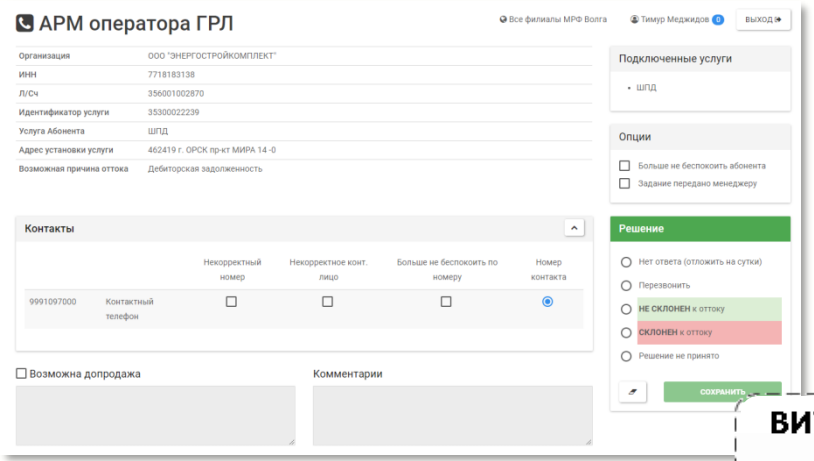


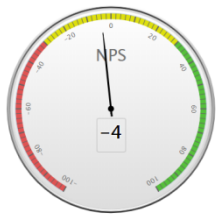
Outflow tendency **86%**

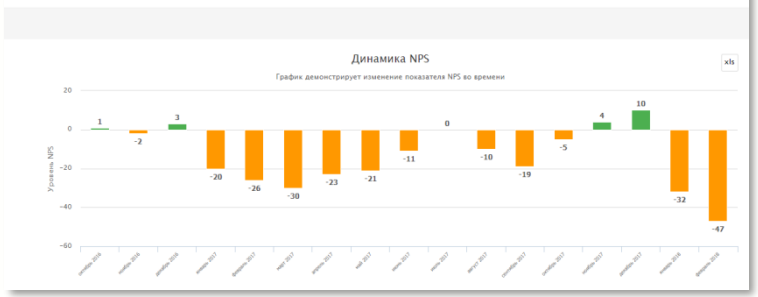
Outflow tendency **78%**

Outflow tendency **65%**

Technological landscape



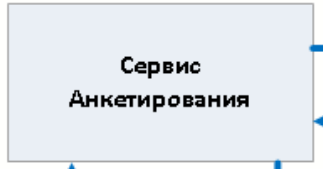




Обогащение данными

Формирование анкеты для клиента

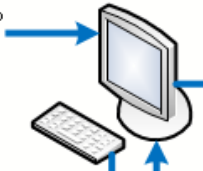
Оценки CSI/NPS



Анкета Email/Телефон

Оценки CSI/NPS

Склонность к оттоку



Звонок

Результат удержания

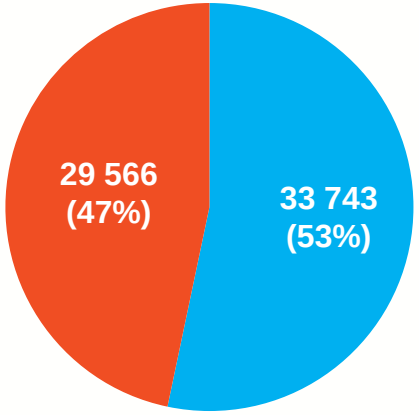
Анкеты

Обратная связь по результатам анкетирования

Key project indicators

Impact on revenue

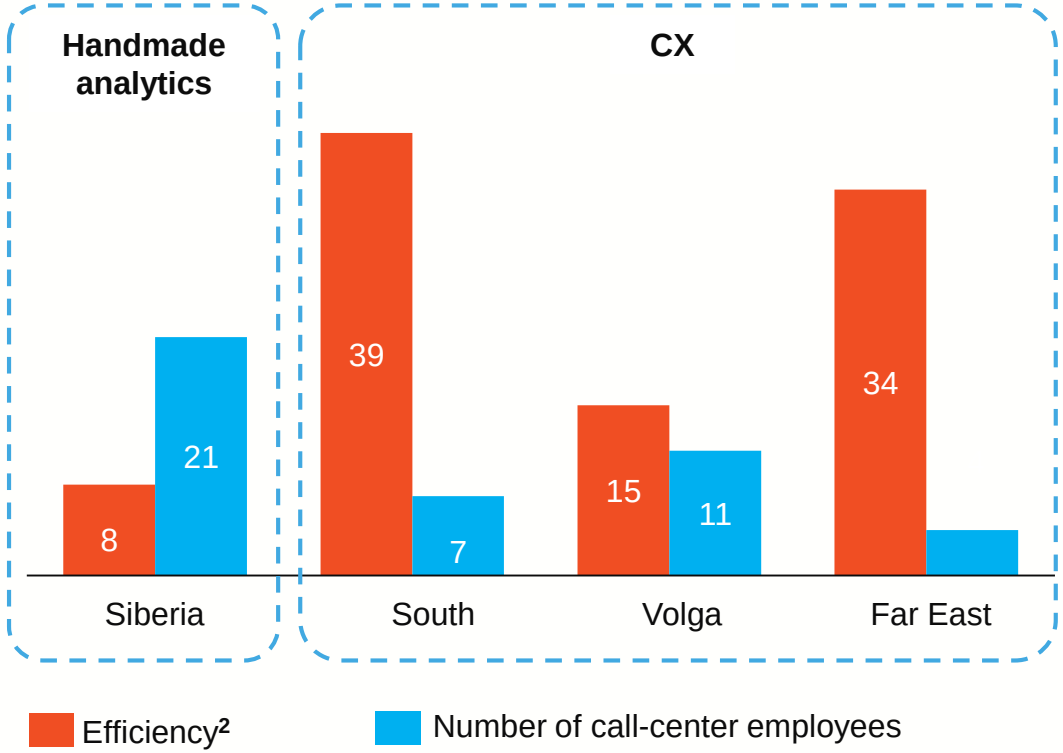
Sample size[1]



State of contact	% gone from sample
Haven't made a call	9.1%
Have made a call / inclined	6.6% (-2,5 пп)

have made a call haven't made a call

Impact on operational efficiency



[1] Period is July 2017 - February 2018

[2] Efficiency.% - percentage of reaching the client that is inclined to outflow to the sample to total quantity of made contacts

Working procedure of Technology forecasting

SUBJECTS

TREND LIFECYCLE

USING SOURCES

COMPLEX MULTIFACTOR ANALYSIS OF TRENDS

SCIENTISTS AND INNOVATORS

INVESTORS

COMMERCIAL COMPANIES

CREATION → DEVELOPEMENT → EXPLOITATION

SIENCE PUBLICATIONS	PATENTS	FINANCIAL INFORMATION (DEALS, BUYS, IPO)	MEDIA, INDUSTRY SITES
4 m of publications	3 m of patents	250 000 companies	100 000 articles
10 000 companies	10 000 companies	100 000 deals	50 sources
100 countries	20 patents bureaus	100 countries	
	100 countries		

COLLECTING DATA TECHNOLOGEIS

DATA PROCESSING AND NORMALIZATION TECHNOLOGEIS

MACHINE LEARNING TECHNOLOGEIS

LIST OF GLOBAL TRENDS AND FACTORS

Trends selection procedure

Science publications

Patents

PRIMARY LIST OF TRENDS

EXPANSION OF TRENDS' LIST USING ANOTHER SOURCES

Investment

MERGE OF SIMILAR TRENDS

Computer vision = Object Recognition + Machine Vision + Image Recognition

Gesture recognition = Sign Language Recognition + Pen-based Interface

SDN = NFV + Openflow

Smart Cities = Urban Computing + Smart Car + Smart Building

Machine Learning = Deep Learning + Support Vector Machine

Deletion of most general trends

Example: Software, Hardware

EXPERT VERIFICATION

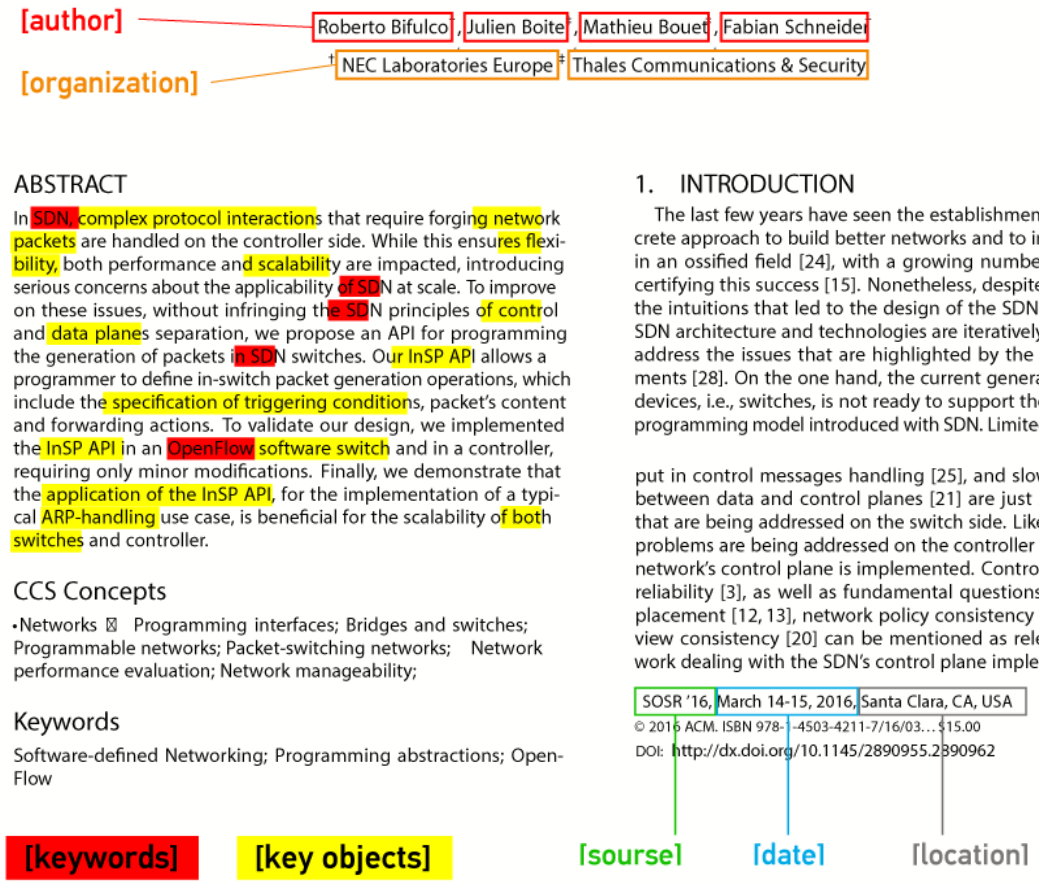
Example:

SDN, Openflow, Software-defined network

EXPANSION WITH SYNONYMS DESCRIPTION OF THE TREND

Technologies of Machine learning

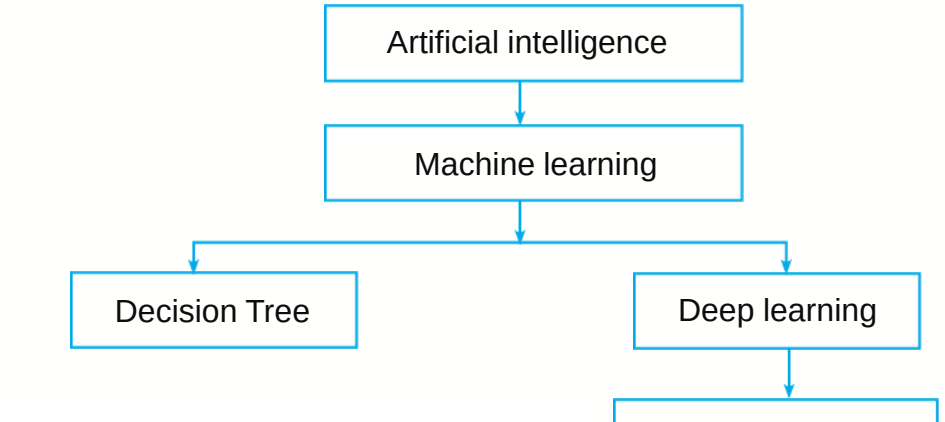
HIGHLIGHTING OF ENTITIES



NORMALIZATION OF DATA

- UCLA
- Univ. of California, LA
- UCLA, Los Angeles, CA
- University of California at Los Angeles

HIERARCHICAL CLASTERING



$$PMI = \log\frac{p(x,y)}{p(x)p(y)} \qquad PM(a|b) = \frac{P(ab)}{P(b)}$$