Информациски системи и големи податоци

Организација на предметот

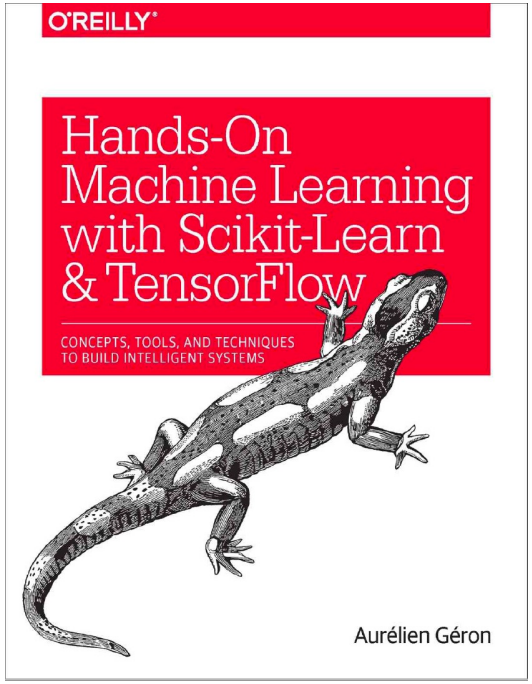
- Предавања и вежби:
  - Вон. Проф. д-р. Христијан Ѓорески
- Контакт: hristijang@feit.ukim.edu.mk
- Консултации – (319) - онлајн,?
- Часови неделно: 2(п) + 2(ав) + 1(лв)
- Полагање (може да се промени)
  - Лабораториски 20%
  - Тест (30%)
  - Семинарска (50%)

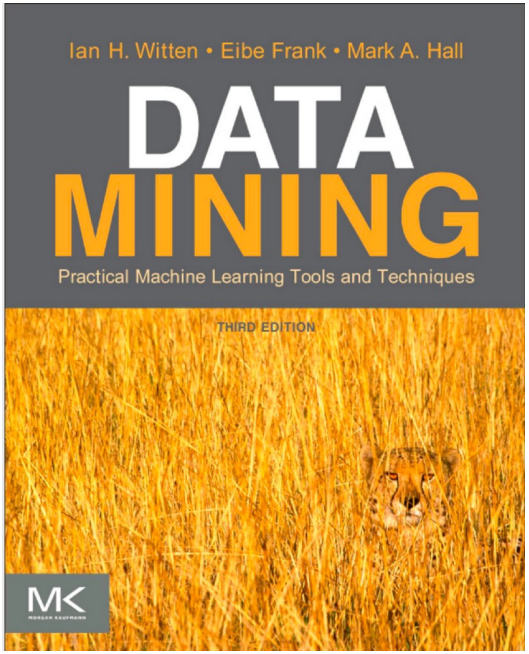
Content

- Data Science & Machine Learning
  - Classification - WEKA
    - Decision Tree, KNN, Naïve Bayes, SVM, Ensembles
    - Evaluation, Performance Metrics
  - Data Processing - Python
    - Data cleaning, outliers, missing values, encoding
  - Regression – WEKA & Python
    - Linear regression, Feature Selection, Hyperparameter Optimization
  - Clustering – WEKA & Python
  - Association Rules - WEKA & Python
- Data Warehouses & Big Data

Ресурси

- Е-курсеви: http://e-kursevi.feit.ukim.edu.mk
- Книги





Софтвер и алатки

- WEKA – Java
- Python
  - Scikit-learn, Pandas, Matplotlib
  - Spyder, Jupyter notebook, Google Colab

Resources used

- Stanford CS229 Machine Learning course lectured by Andrew Ng
- JS IPS – course Data Mining
- University of Pennsylvania - CIS 419/519 - Introduction to Machine Learning
- https://machinelearningmastery.com

Widening Research on Pervasive and eHealth

WideHealth

Goals

- To enable a **new generation of researchers** in the widening countries to **develop and adapt novel eHealth technologies**.
- **Establish a sustainable network** of knowledge research and dissemination across Europe in the **Pervasive Health topic**

- Seminars, Workshops, Newsletters, eHealth Materials

https://widehealth.eu

Follow us on

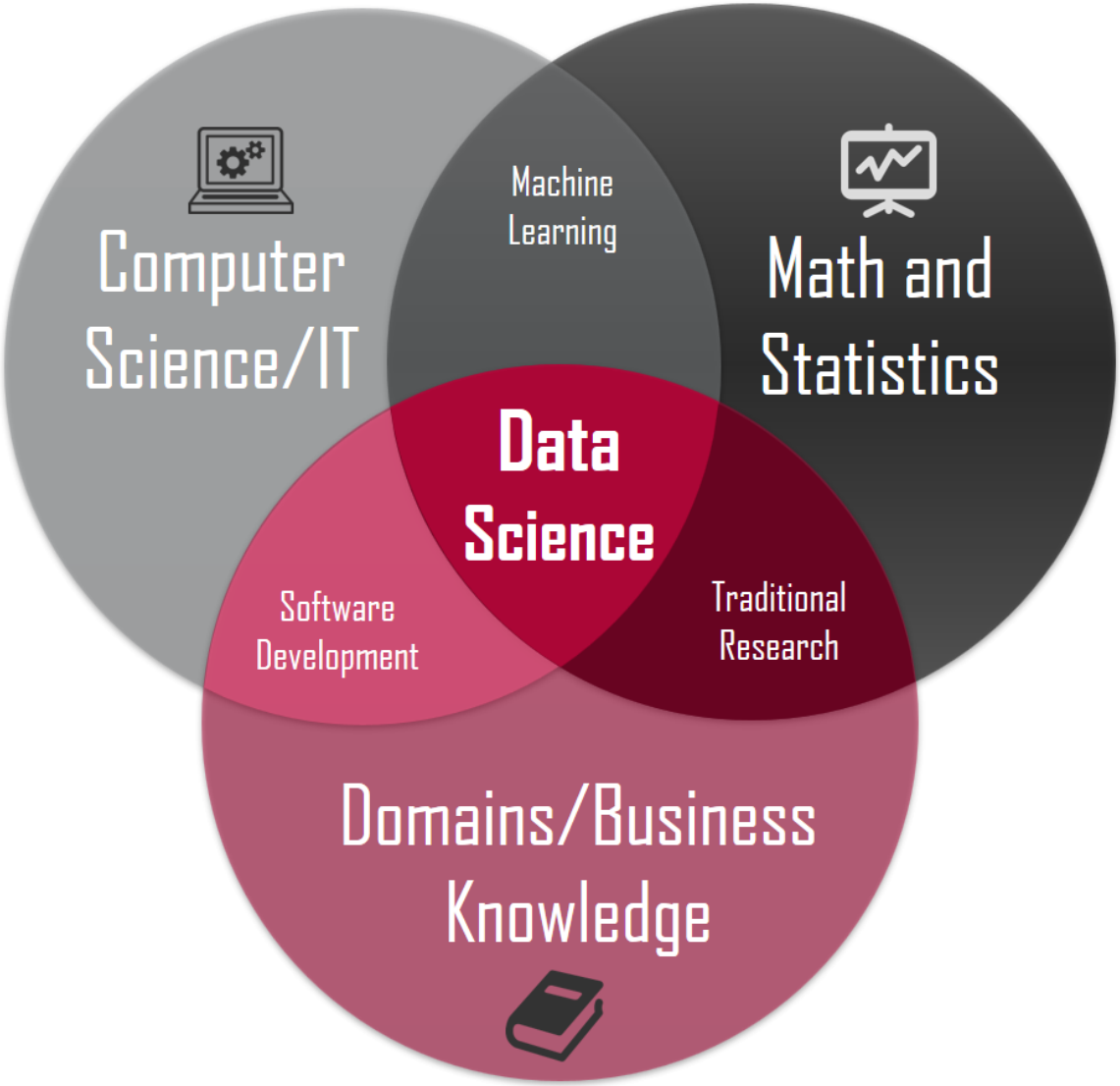
https://www.linkedin.com/in/widehealth-project-eu-105610207/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952279"

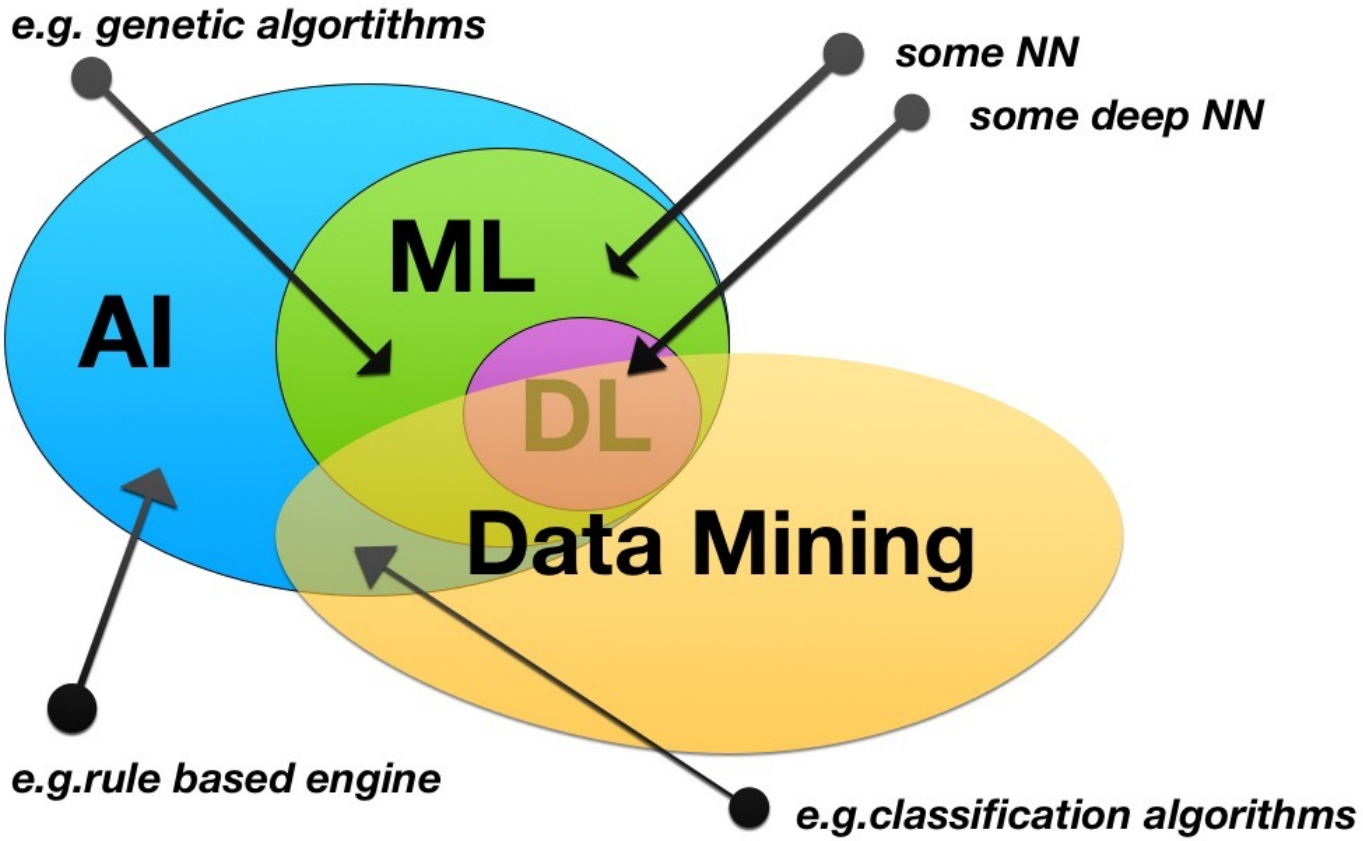
Data Science Macedonia

https://www.meetup.com/Data-Science-Macedonia/

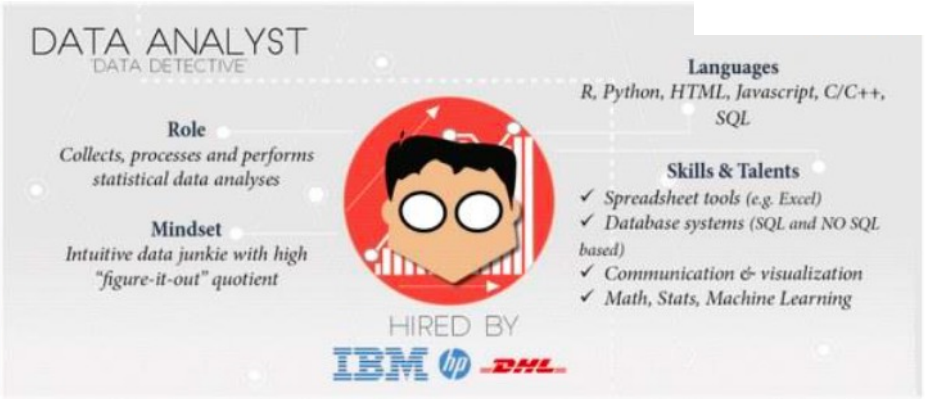
Perspective

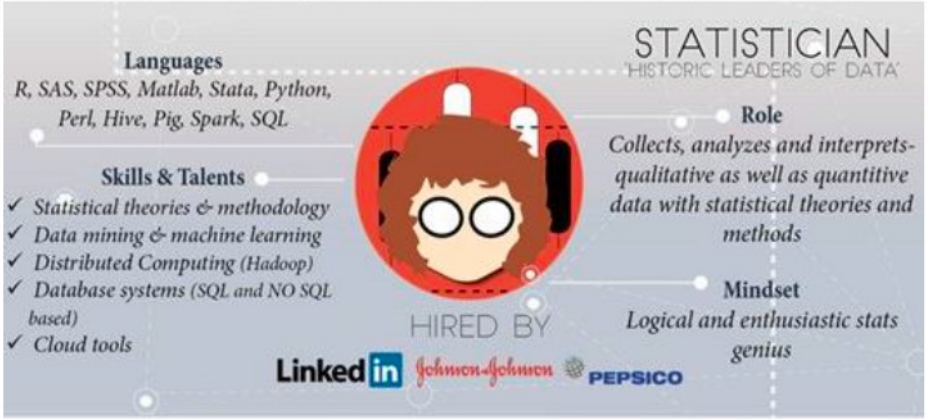


Research Perspective



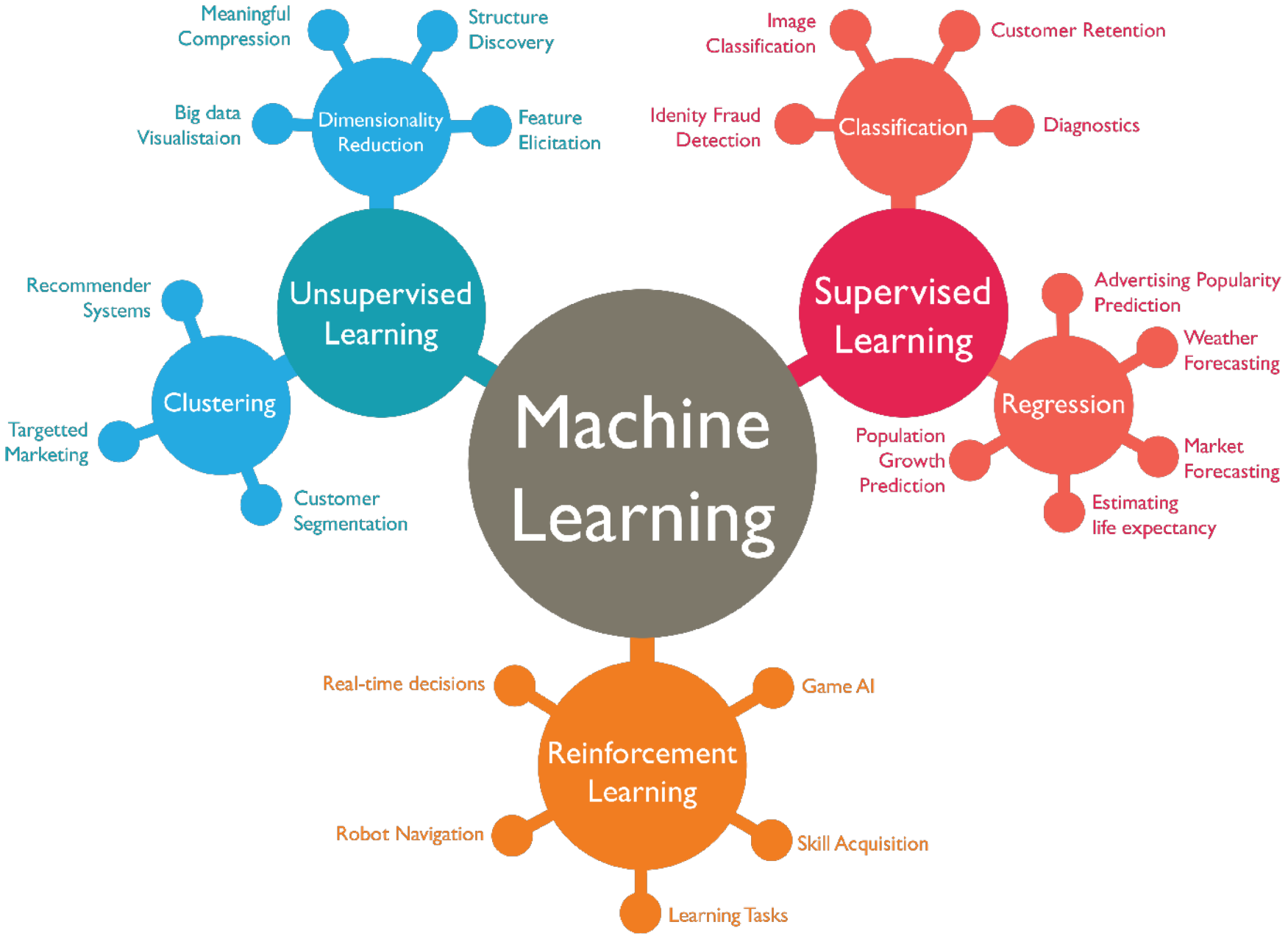
Data Scientists? - Unicorn





$$$$$

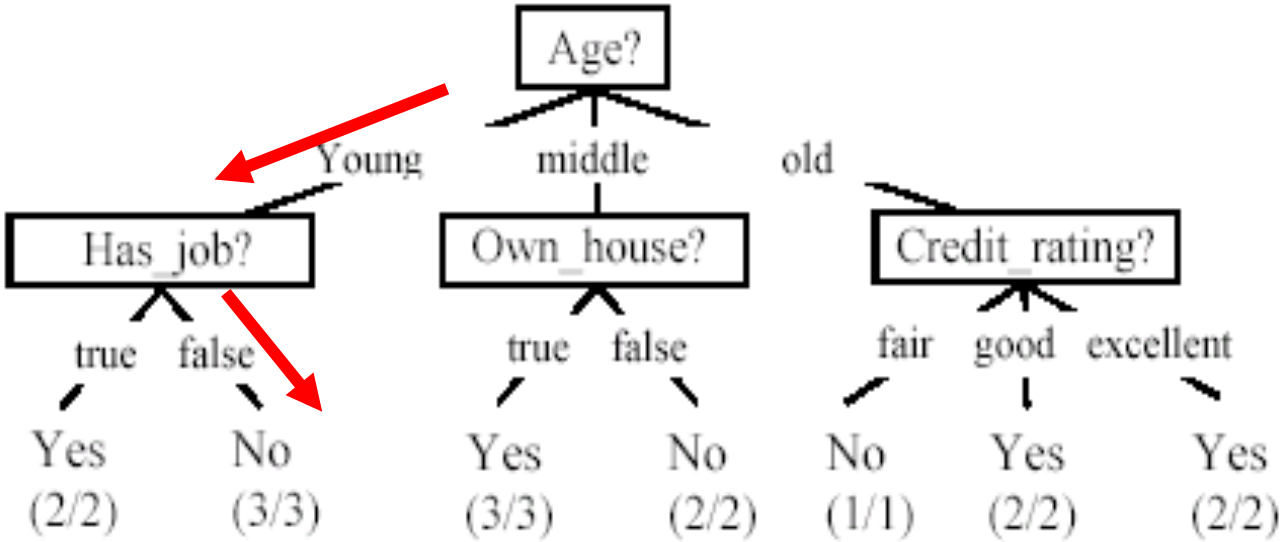
$$$


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating life expectancy
Population Growth Prediction
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Use the decision tree

Age	Has_Job	Own_house	Credit-Rating	Class
young	false	false	good	?

No



A Brief History of Machine Learning

History of Machine Learning

- 1950s
  - Samuel's checker player
  - Selfridge's Pandemonium
- 1960s:
  - Neural networks: Perceptron
  - Pattern recognition
  - Learning in the limit theory
  - Minsky and Papert prove limitations of Perceptron
- 1970s:
  - Symbolic concept induction
  - Winston's arch learner
  - Expert systems and the knowledge acquisition bottleneck
  - Quinlan's ID3
  - Michalski's AQ and soybean diagnosis
  - Scientific discovery with BACON
  - Mathematical discovery with AM

Slide credit: Ray Mooney

History of Machine Learning (cont.)

- 1980s:
  - Advanced decision tree and rule learning
  - Explanation-based Learning (EBL)
  - Learning and planning and problem solving
  - Utility problem
  - Analogy
  - Cognitive architectures
  - Resurgence of neural networks (connectionism, backpropagation)
  - Valiant's PAC Learning Theory
  - Focus on experimental methodology
- 1990s
  - Data mining
  - Adaptive software agents and web applications
  - Text learning
  - Reinforcement learning (RL)
  - Inductive Logic Programming (ILP)
  - Ensembles: Bagging, Boosting, and Stacking
  - Bayes Net learning

Slide credit: Ray Mooney

History of Machine Learning (cont.)

- 2000s
  - Support vector machines & kernel methods
  - Graphical models
  - Statistical relational learning
  - Transfer learning
  - Sequence labeling
  - Collective classification and structured outputs
  - Computer Systems Applications (Compilers, Debugging, Graphics, Security)
  - E-mail management
  - Personalized assistants that learn
  - Learning in robotics and vision
- 2010s
  - Deep learning systems
  - Learning for big data
  - Bayesian methods
  - Multi-task & lifelong learning
  - Applications to vision, speech, social networks, learning to read, etc.

Based on slide by Ray Mooney