Welcome to Machine Learning Module

Huntsman Faculty & Staff Directory

Economics and Finance Department

Management Department

Management Information Systems Department

Marketing and Strategy Department

School of Accountancy

Pedram Jahangiry

Visiting Assistant Professor

Department(s):
Economics and Finance Department

Educational Background

PhD, Economics, Arizona State University, 2017
Master, Economics, Simon Fraser University, 2013
MBA, Sharif University, 2012
Industrial Engineering, IUST, 2009

Biography

Pedram Jahangiry, PhD is an assistant professor in the Economics and Finance Department of the Jon M. Huntsman School of Business at Utah State University. He recently passed level III in the CFA program. Prior to joining the Huntsman School in 2018, Pedram was a research associate within Financial Modeling Group at BlackRock NYC. His research is involved in machine learning applications in finance, empirical asset pricing, and factor models.
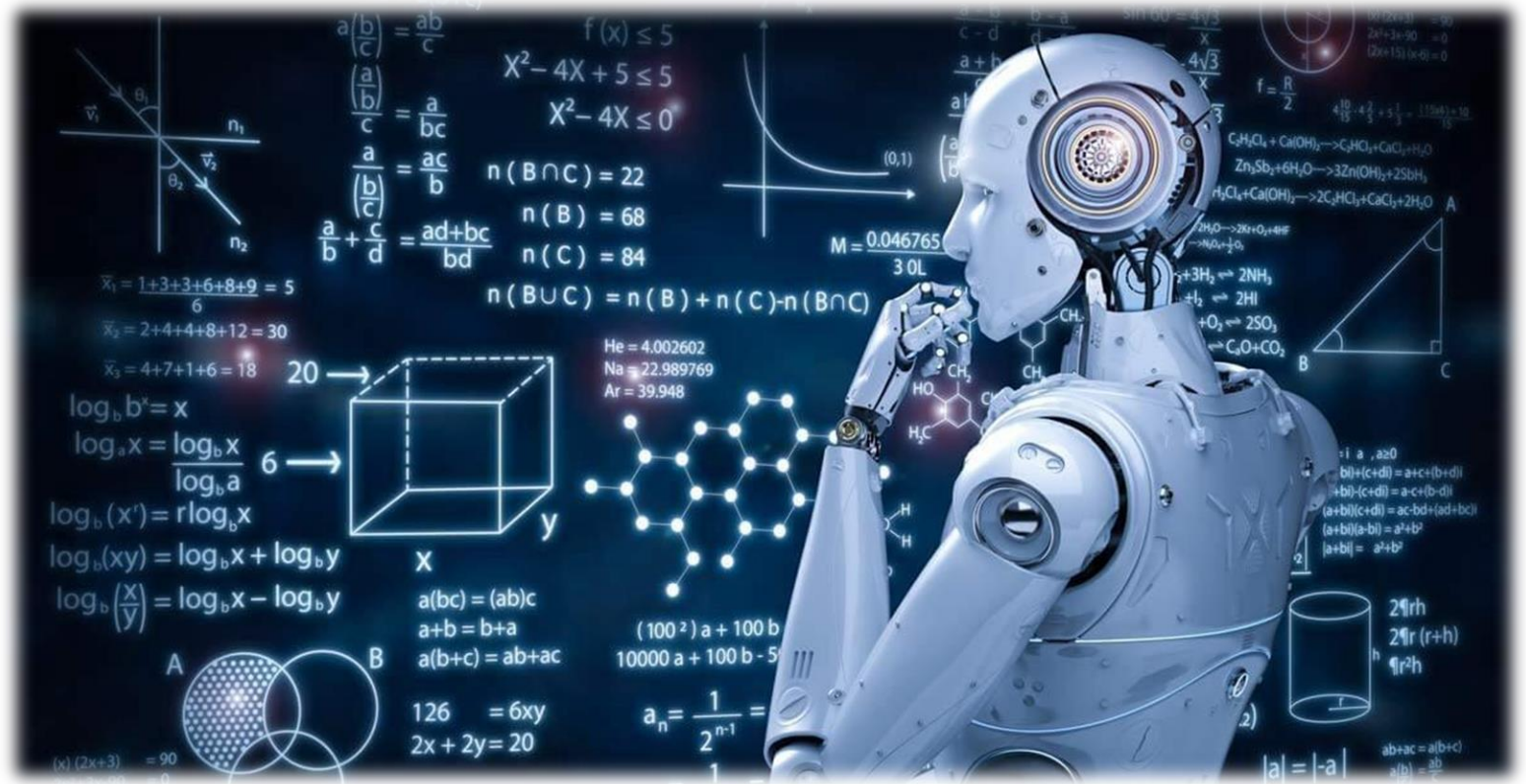
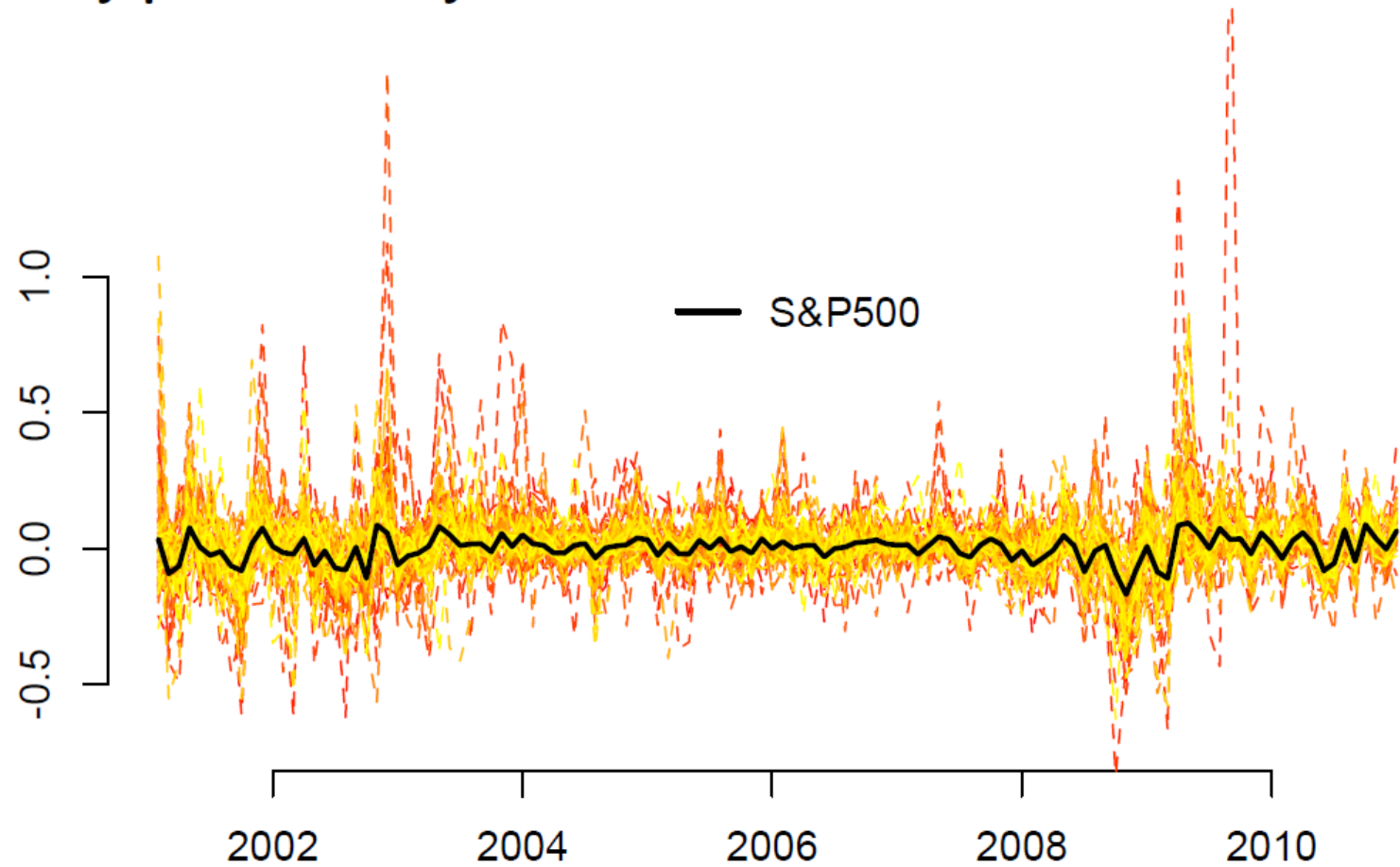Show more

Contact Information

Office Location:
Eccles Business Building 507

Phone: 435.797.2345

Email: pedram.jahangiry@usu.edu

Personal Website   **Curriculum Vitae**

- What is Machine Learning?
- Different major types
- Applications
- Key terminology
- Applied ML
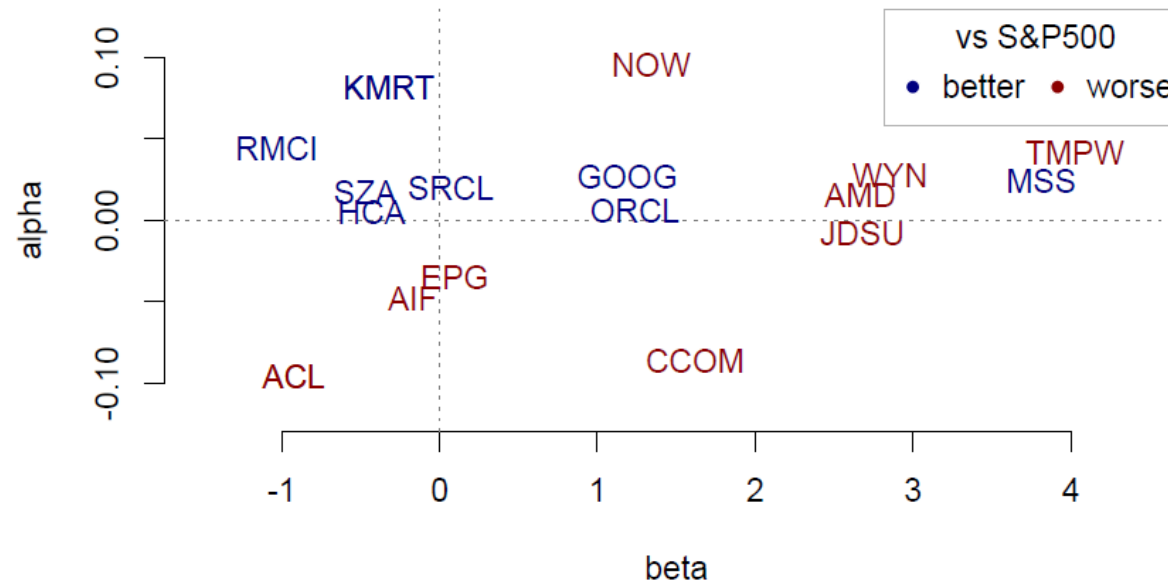
Fancy plot: monthly stock returns

What do we learn?

Useful plot: market model coefficients

Fit a line between stock returns R_t and market returns M_t (SP).
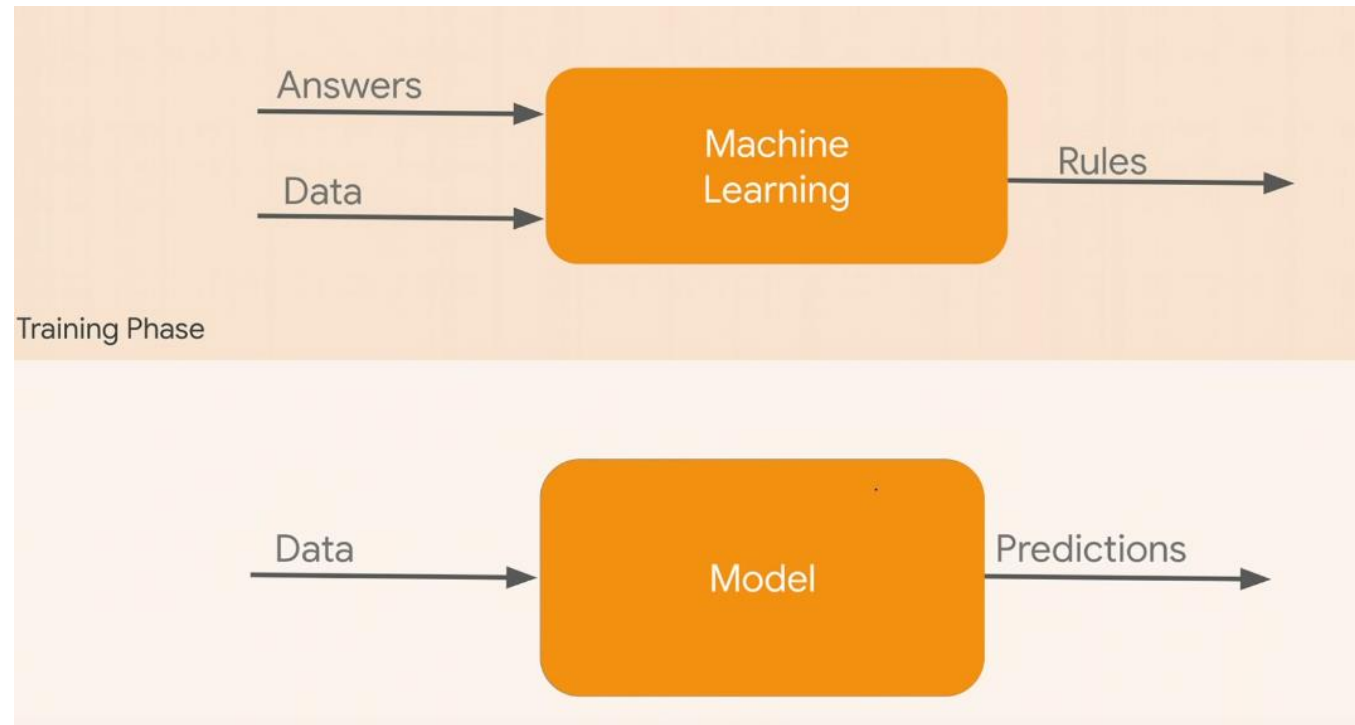
$$R_t \approx \alpha + \beta M_t$$

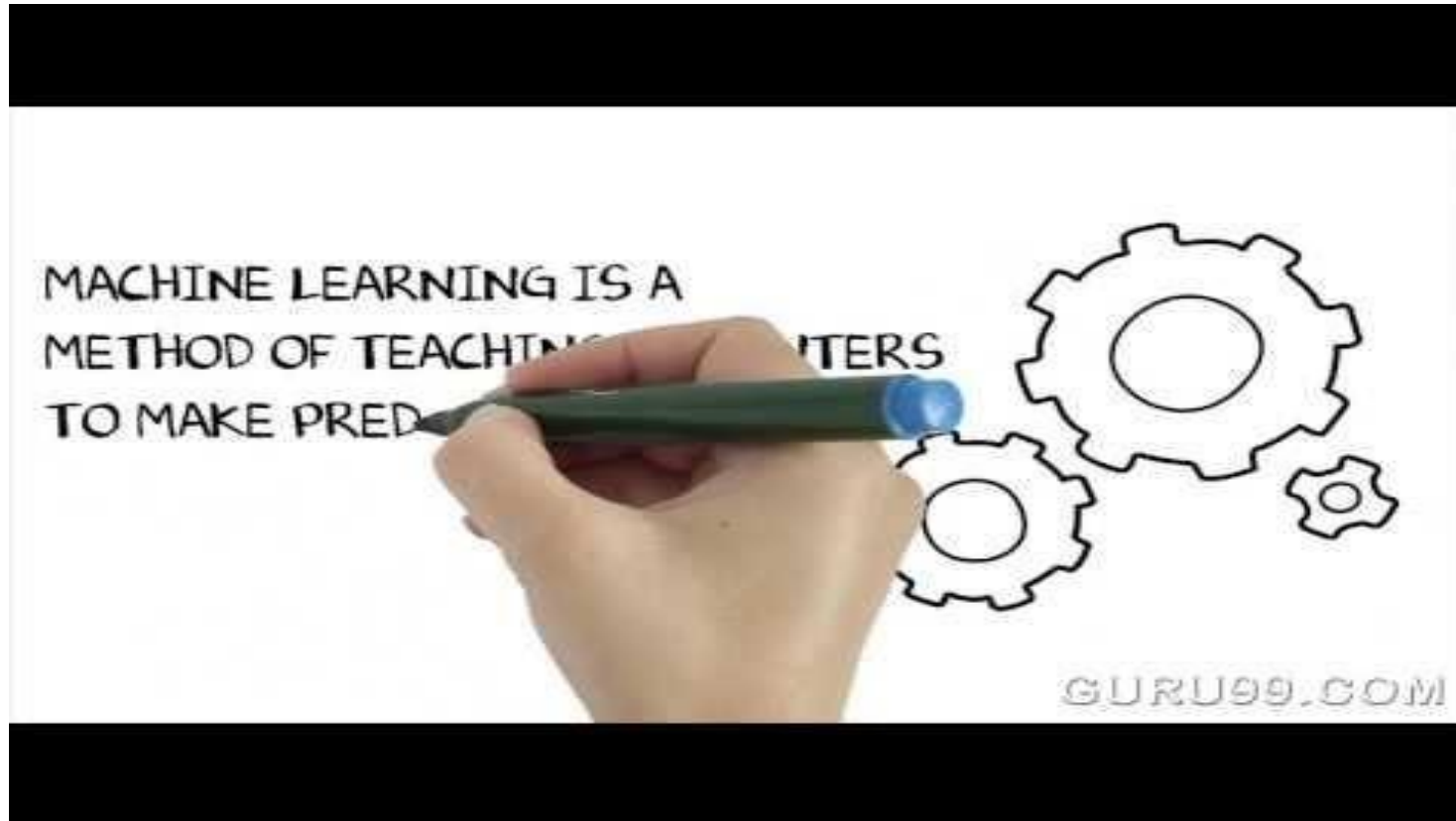α is money you make regardless of what the market does.
β is the asset's sensitivity to broad market movements.

Machine Learning vs. Traditional Programming

What is Machine Learning?

What is Machine Learning?

Machine Learning is the filed that thinks about how to **automatically** build robust **predictions** from **complex data**.

- ML is closely related to modern statistics (many ideas come from statisticians!)
- Statistics focus on model inference, ML focus on predictive performance
- The entire filed of ML is calibrated against "**out-of-sample**" experiments that evaluate how well a model **trained** on one dataset (train data) will **predict** new data (test data).
- While there is a recent push to build more transparency into machine learning, wise practitioners avoid assigning **structural** meaning to the parameters of their fitted models.
- ML models are **black boxes** whose purpose is to do a good job in **predicting** a future that follows the **same patterns** as in past data.

Why you need to learn ML?

- To project the information in your data into a **low-dimensional space** that contains key insights for the decisions you need to make

- To give you the tools to quickly **turn messy data into useful information** that has direct relevance to business policy

- Big data + Machine learning = Data science
- Big data + Machine learning + Economics/Econometrics= **Business data science**

Understanding the major types of machine learning

Supervised learning
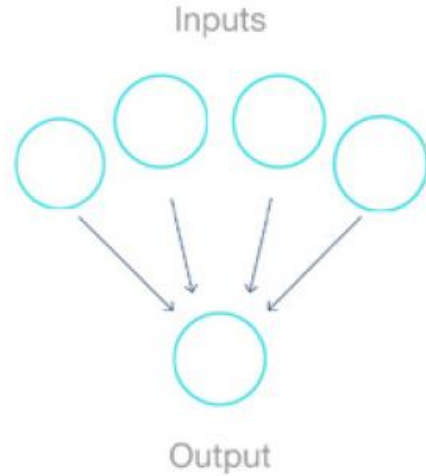
Inputs

Output

Unsupervised learning

Reinforcement learning

Algorithm

Reward

State

Action

Environment
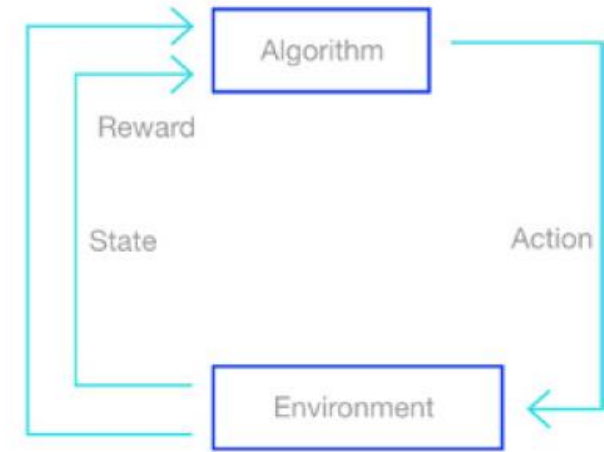
What it is

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output (eg, how the inputs "time of year" and "interest rates" predict housing prices)

An algorithm explores input data without being given an explicit output variable (eg, explores customer demographic data to identify patterns)

An algorithm learns to perform a task simply by trying to maximize rewards it receives for its actions (eg, maximizes points it receives for increasing returns of an investment portfolio)

Machine Learning Branches
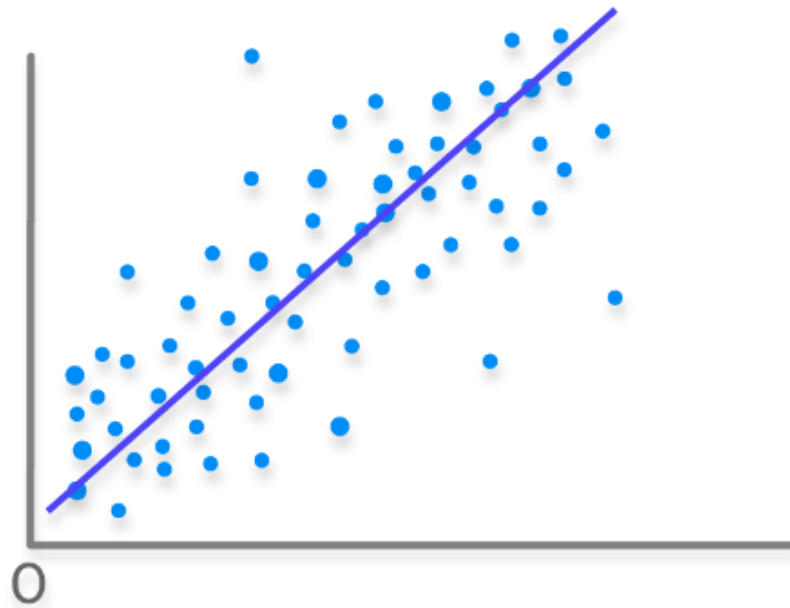
Regression

Regression models (both linear and non-linear) are used for predicting a real value, like salary for example. If your independent variable is time, then you are forecasting future values, otherwise – your model is predicting present but unknown values. Regression techniques vary from MLR to SVR and Boosted Trees.

5 Applications
Of Regression in Real world

DatumGuy

Future Prediction

To predict events yet to occur. For example predicting future sales, units consumed by consumer.

Process Optimization

To optimize business processes like a manager in call center would like to know how much call waiting time affects the number of complaints.

Market Forecast
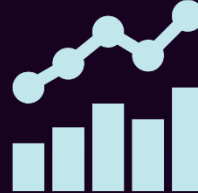
Here Multivariate linear technique is used to forecast sales volumes, or market movement.

Human Resources

Linear regression in also used to predict the demographics and types of future work force in large companies . This way companies can create better hiring plans.

Study of Environment

Linear Regression have a wide applications in Environment Science. For example in Canada, statistical analysis is used to measure effect of oil spills on aquatic ecosystem.

Machine Learning Branches

Unlike regression where you predict a continuous number, you use classification to predict a category. There is a wide variety of classification applications from medicine to marketing. Classification models include linear models like Logistic Regression, SVM, and nonlinear ones like K-NN, Kernel SVM and Random Forests.

Classification

Machine Learning Branches
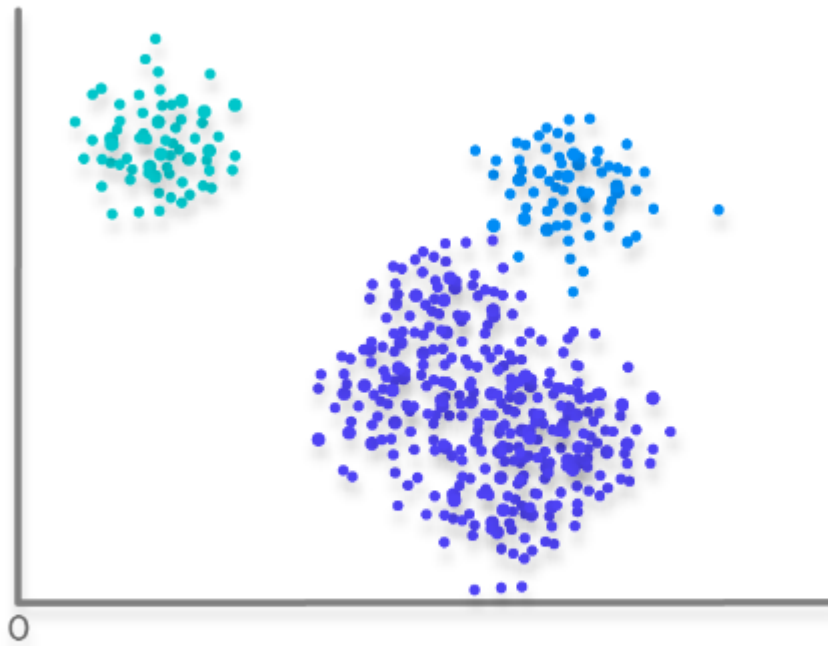
Clustering

Clustering is similar to classification, but the basis is different – in Clustering you don't know what you are looking for. When you use clustering algorithms on your dataset, unexpected things can suddenly pop up – like structures, clusters and groupings you would have never thought of otherwise.

5 Applications Of Clustering In Real World

Segmenting Customers

Used in marketing, where customers are are subdivided into segments with each segment consisting of customers with similar characteristics.

Creating Newsfeeds

To cluster articles from various sources by their similarity — it can separate documents into disjoint clusters and hence helps decision making.

Pattern Recognition

This feature is readily used in medical field these days where algorithm automatically detect for segmentation of blood cells for leukemia detection or for cancer cells detection.

Cloud Storage

To enhance performance, capacity, or reliability , clustering distributes work loads to each server, and provides access to all files from any server regardless of the physical location of the file.

Recommender System

To recommend similar items for customers. To find similar items, comparison b/w the set of users who like each item is done, if a similar set of users like two different items, then the items themselves are probably similar!

DatumGuy

Machine Learning Branches

Reinforcement learning algorithms include techniques like Thompson Sampling, Upper Confidence Bound and Q-Learning. These are used a lot when training machines to perform tasks such as walking. Desired outcomes provide the AI with reward, undesired – with punishment. Machines learn through trial and error.

Reinforcement Learning

Machine Learning Branches

Natural Language Processing

Teaching machines to understand what is said in spoken and written word is the focus of Natural Language Processing. Whenever you dictate something into your iPhone / Android device and it's converted to text – that's an NLP algorithm in action. Methods include decision trees, Markov processes, and more.
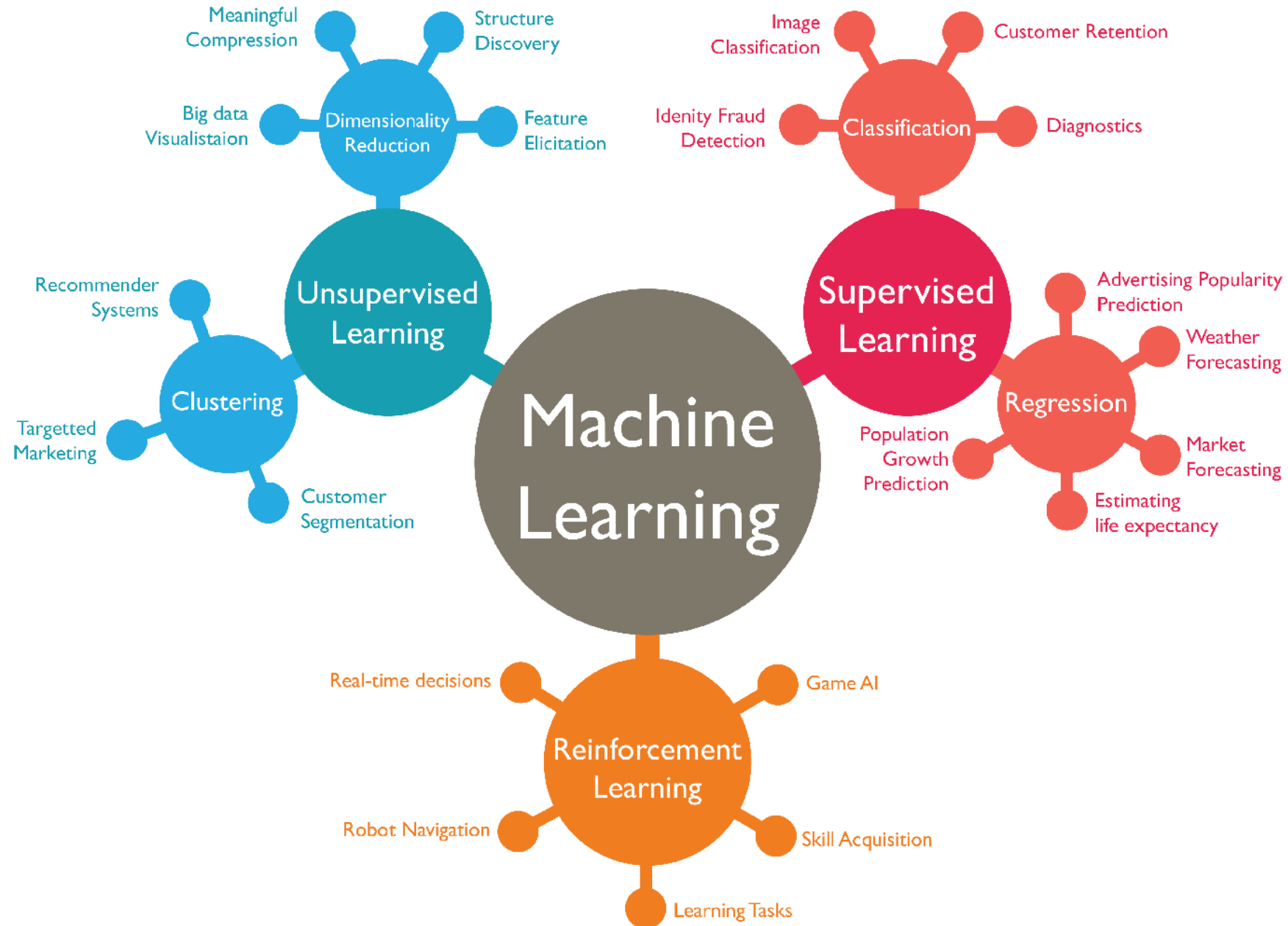
- Familiarize you with ML terminology
- Prepare you for job interviews
- Provide cheat sheets and resources
- Help you to hedge yourself against the next recession

Hedge yourself against the next recession!

When Will Next U.S. Recession Begin?

Wall Street Journal survey of private-sector economists

Pulsenomics survey of economists, investment strategists and real estate market analysts

Year	WSJ	Pulsenomics
2018	0%	4%
2019	8%	24%
2020	59%	48%
2021	22%	14%
2022	8%	9%
after 2022	4%	1%

Keeping Current Matters

Pulsenomics & WSJ

INDUSTRY
Machine Learning

Accommodation & Food	Agriculture	Banking & Insurance
Biotechnological & Life Sciences	Construction & Engineering	Education & Research
Emergency & Relief	Finance	Manufacturing
Government and Public Works	Healthcare	Media & Publishing
Justice, Law and Regulations	Miscellaneous	Accounting
Real Estate, Rental & Leasing	Utilities	Wholesale & Retail

- Accommodation & Food
  - Food
  - Restaurant
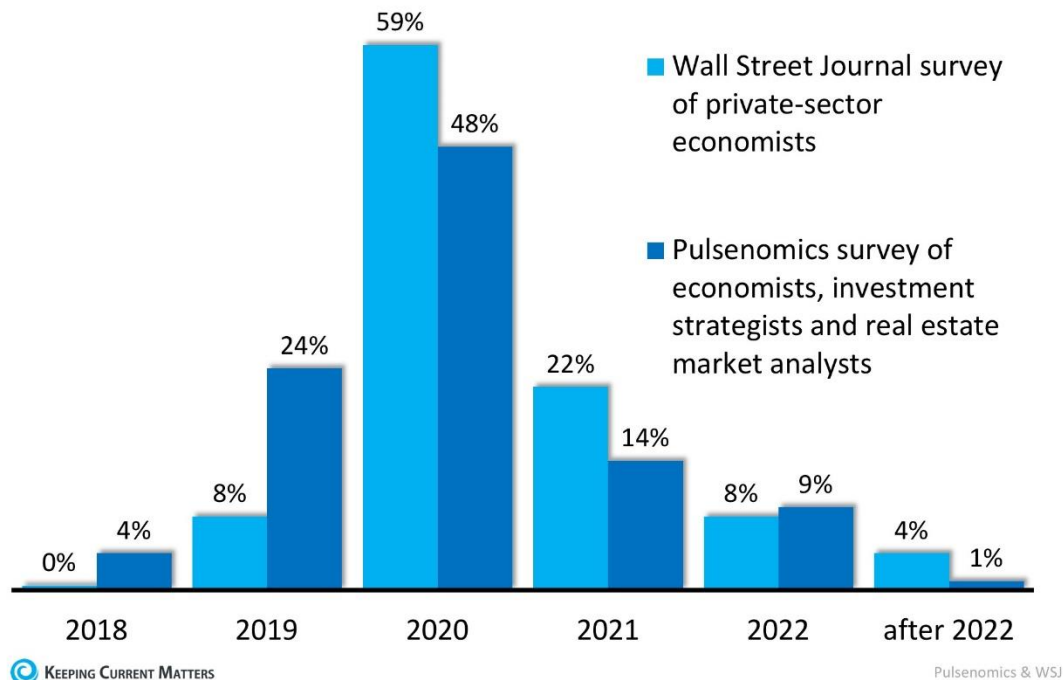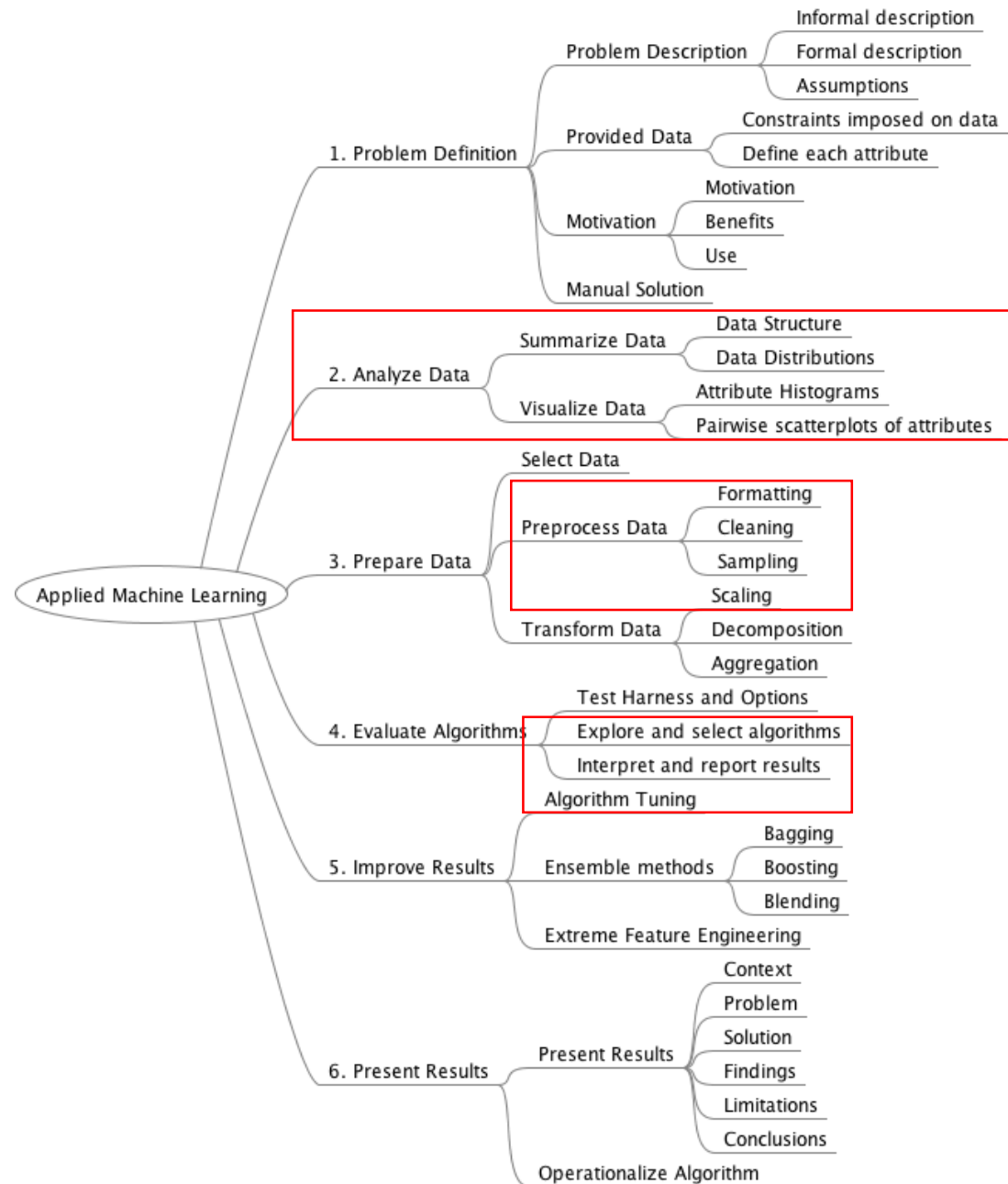 - Accommodation
- Accounting
  - Machine Learning
  - Analytics
  - Textual Analysis
  - Data
  - Research and Articles
  - Websites
  - Courses
- Agriculture
  - Economics
  - Development
- Banking & Insurance
  - Consumer Financial
  - Management and Operations
  - Valuation
  - Fraud
  - Insurance and Risk
  - Physical

- Biotechnological & Life Sciences
  - General
  - Sequencing
  - Chemoinformatics and drug discovery
  - Genomics
  - Life-sciences
- Construction & Engineering
  - Construction
  - Engineering
  - Material Science
- Economics
  - General
  - Machine Learning
  - Computational
- Education & Research
  - Student
  - School
- Emergency & Relief
  - Preventative and Reactive
  - Crime
  - Ambulance
  - Disaster Management

- Finance
  - Trading & Investment
  - Data
- Healthcare
  - General
- Justice, Law and Regulations
  - Tools
  - Policy and Regulatory
  - Judicial
- Manufacturing
  - General
  - Maintenance
  - Failure
  - Quality
- Media & Publishing
  - Marketing
- Miscellaneous
  - Art
  - Tourism
- Physics
  - General
  - Machine Learning

- Physics
  - General
  - Machine Learning
- Government and Public Works
  - Social Policies
  - Election Analysis
  - Disaster Management
  - Politics
  - Charities
- Real Estate, Rental & Leasing
  - Real Estate
  - Rental & Leasing
- Utilities
  - Electricity
  - Coal, Oil & Gas
  - Water & Pollution
  - Transportation
- Wholesale & Retail
  - Wholesale
  - Retail

What we learn
in this class?

Applied Machine Learning

1. Problem Definition
- Problem Description
  - Informal description
  - Formal description
  - Assumptions
- Provided Data
  - Constraints imposed on data
  - Define each attribute
- Motivation
  - Motivation
  - Benefits
  - Use
- Manual Solution

2. Analyze Data
- Summarize Data
  - Data Structure
  - Data Distributions
- Visualize Data
  - Attribute Histograms
  - Pairwise scatterplots of attributes

3. Prepare Data
- Select Data
- Preprocess Data
  - Formatting
  - Cleaning
  - Sampling
- Scaling
- Transform Data
  - Decomposition
  - Aggregation

4. Evaluate Algorithms
- Test Harness and Options
- Explore and select algorithms
- Interpret and report results
- Algorithm Tuning

5. Improve Results
- Ensemble methods
  - Bagging
  - Boosting
  - Blending
- Extreme Feature Engineering

6. Present Results
- Present Results
  - Context
  - Problem
  - Solution
  - Findings
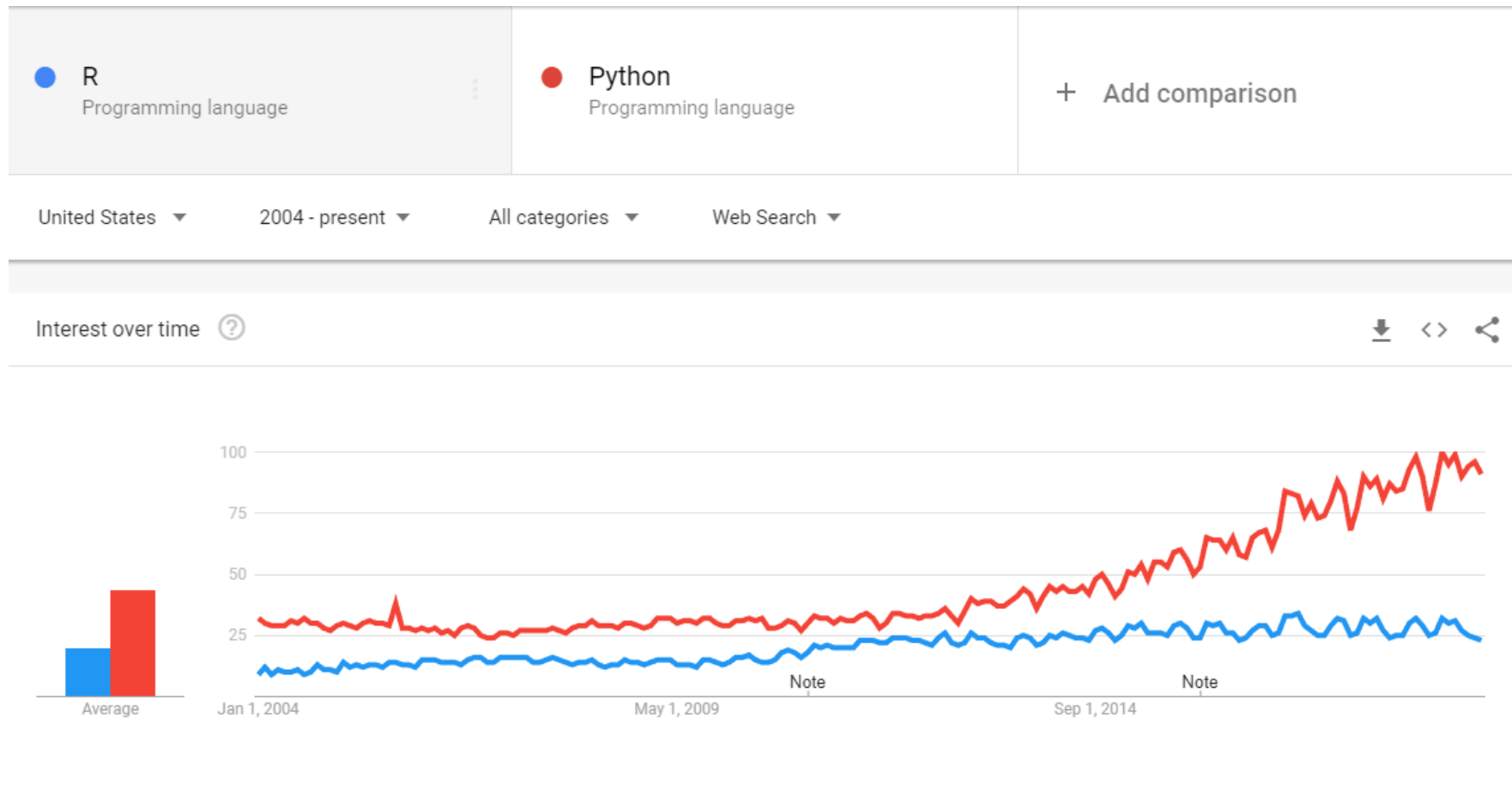 - Limitations
  - Conclusions
- Operationalize Algorithm

Classes and events

Why Python?

TOP PYTHON / SCALA / R LIBRARIES IN DATA SCIENCE

PYTHON	COMMITS	CONTRIBUTORS		SCALA	COMMITS	CONTRIBUTORS		R	COMMITS	CONTRIBUTORS
TensorFlow	34 689	1 505		DEEPLEARNING4J	22 938	223		H$_2$O	22 947	104
scikit learn	22 811	1 099		Spark MLlib & ML	22 073	1 241		mlr	4 052	59
PYTORCH	11 763	668		PredictionIO	4 383	125		dmlc XGBoost	3 296	286
Keras	4 591	688		bigDL	2 358	50		caret	1 741	65
XGBoost	3 296	286		summingbird	1 780	33		LightGBM	1 125	82
CatBoost	1 746	63		DeepLearning.scala	1 874	15		Prophet	284	34
LightGBM	1 125	82		SPARKLING WATER	1 416	33		ggplot2	4 070	152
dist-keras	1 125	5		Smile	1 195	30		plotly	3 185	24
elephas	170	13		Conjecture	155	8		ggvis	2 159	21
spark-deep-learning	68	11		VEGAS	215	16		DT DataTables	1 921	22
eli5	929	6		Breeze-viz	29	3		rCharts	638	11
matplotlib	25 984	727		akka	22 298	552		Corrplot	335	10
Bokeh	17 066	298		kafka	5 145	442		lattice	781	3
plotly	2 910	47		Breeze	3 431	89		dplyr	4 990	151
Seaborn	2 093	85		Epic	1 790	15		data.table	3 416	50
pydot	169	12		Puck	536	1		lubridate	1 501	51
SciPy	19 255	613		Saddle	184	10		jsonlite	914	12
NumPy	18 162	654		Scalalab	44	1		knitr	5 528	105
Natural Language ToolKit	13 053	238		Slick	2 037	107		rmarkdown	2 373	64
StatsModels Statistics in Python	10 244	153		ADAM	1 775	64		Slidify	302	6
spaCy	8 655	226								
gensim	3 610	275								
pandas	17 243	1 179								
Scrapy	6 643	285								

LEGEND: ● Machine Learning ● Visualization ● Mathematics & Engineering ● Data Manipulation & Analysis ● Reproducible Research

Introduce your self!

What project are you working on?