iAdvize

dictanova

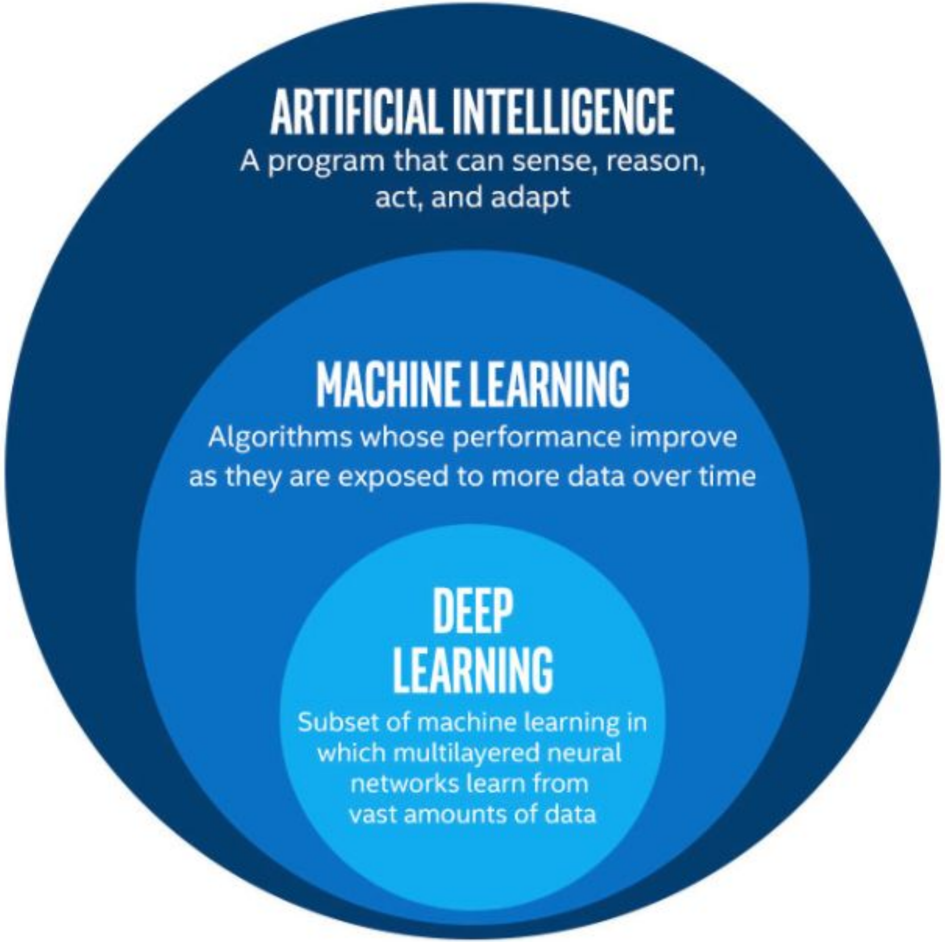
CogniTalk

{EPITECH.
L'ECOLE DE L'INNOVATION ET DE
L'EXPERTISE INFORMATIQUE}

ML WEEK

THE MACHINE LEARNING WEEK

DETECTION
BIG DATA
COMPUTING
INFORMATION TECHNOLOGY
SOCIAL MEDIA
WWW
PROMOTION
SERVICES
MODELS
E-MARKETING
COMMUNICATION
CONTENT
PROCESSING
MULTIMEDIA
NETWORK
PROJECTS
BRANDING
COMPUTER
CONSUMER
SOFTWARE
CONSUMER
DEMAND
MARKETS
ORGANIZATION
CODING
WEB MARKETING
PREDICTIVE
PLANNING
PROGRAM
ANALYTICS
EVENTS
DATA MINING
PROGRAMMING
DATA SCIENCE
INFORMATION
MACHINE LEARNING
VISION
ENGINEERING
RESEARCH
PROBABILITY
KDD
WEB SERVICES
PROJECTS
BIG
SERVICE
WEB DEV
COMPUTING
STRATEGY
WORLDWIDE
MATHS
PATTERN
DATA
VISUALIZATION
DATA
BIG
NETWORK
ENGINEERING
SOLUTIONS
PRO
CODING
PRICING
SOCIAL
PLANNING
MEDIA
MOBILE
SEGMENTATION
STATISTICS
INFORMATION
SOCIAL NETWORKS
TARGET
DIGITAL



ARTIFICIAL INTELLIGENCE

A program that can sense, reason, act, and adapt

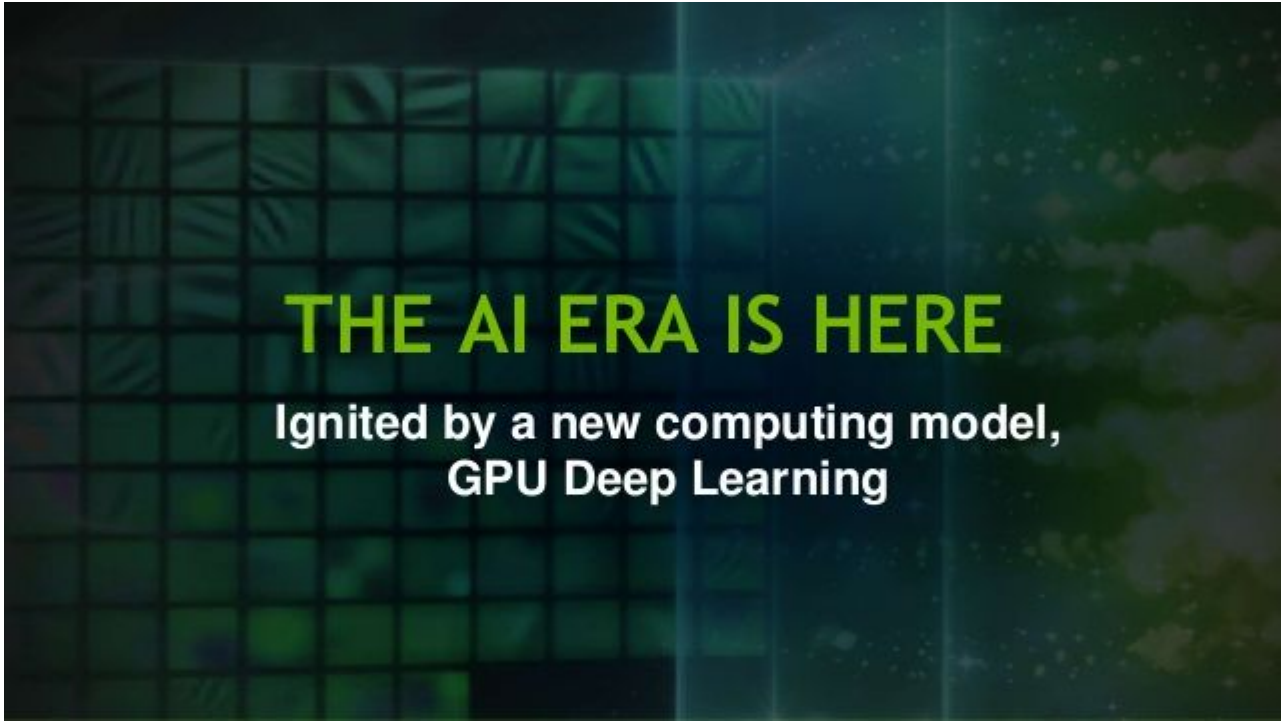
MACHINE LEARNING

Algorithms whose performance improve as they are exposed to more data over time

DEEP LEARNING

Subset of machine learning in which multilayered neural networks learn from vast amounts of data

http://www.prowesscorp.com/whats-the-difference-between-artificial-intelligence-ai-machine-learning-and-deep-learning/

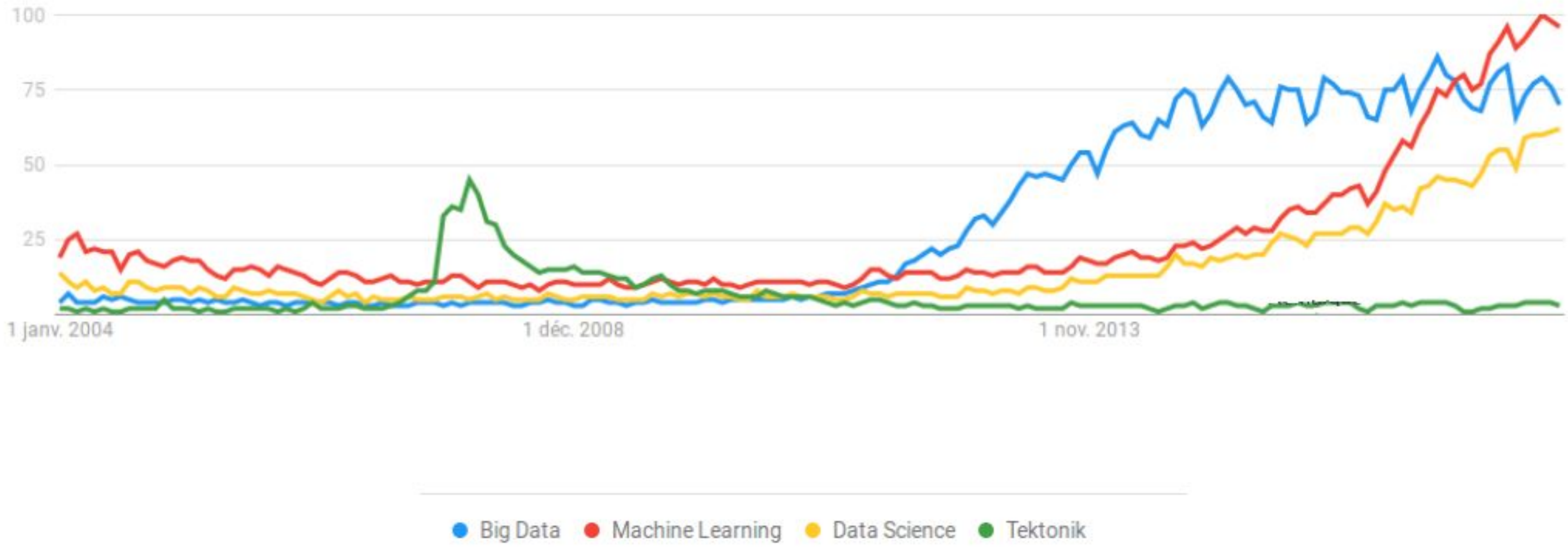


THE AI ERA IS HERE

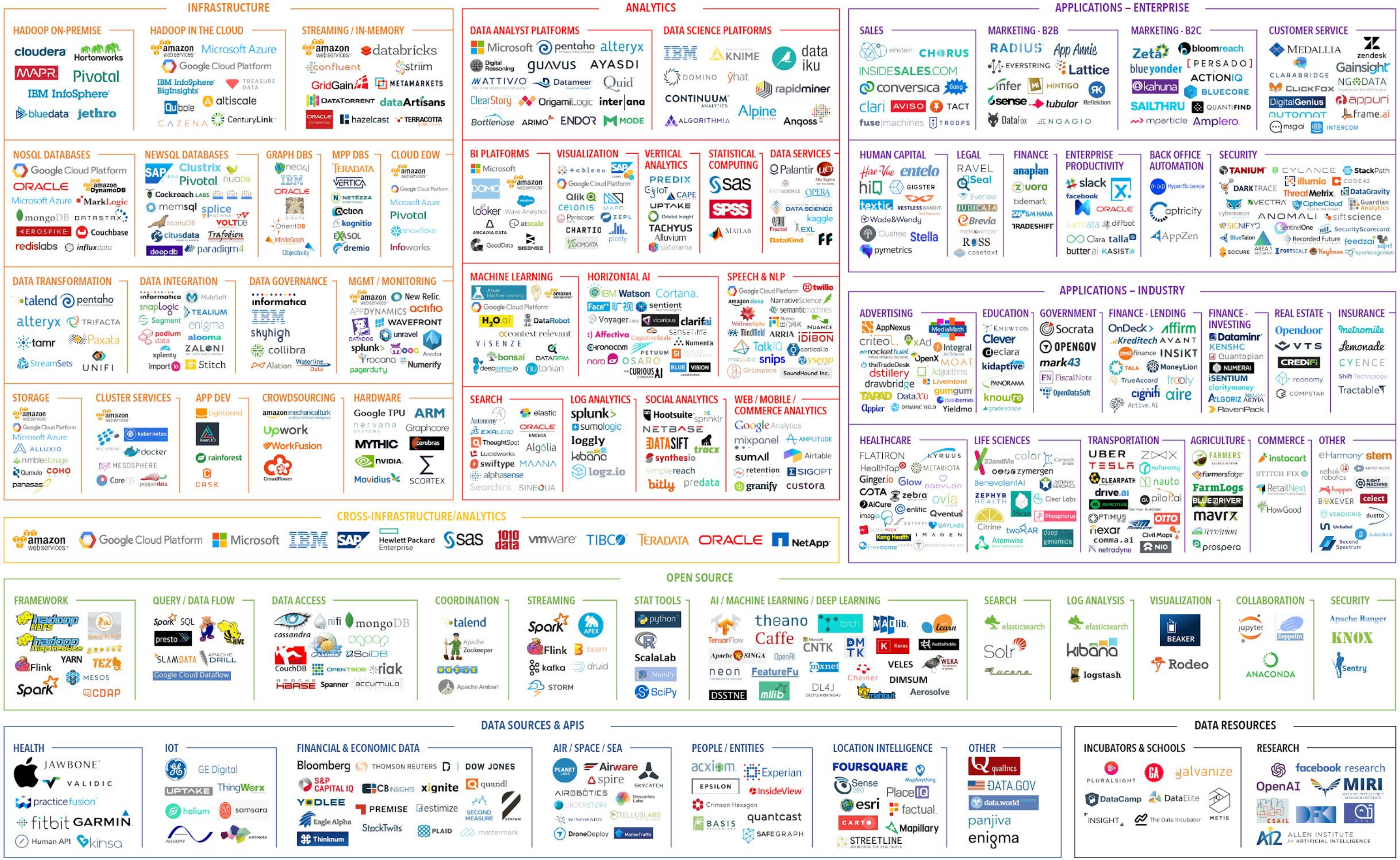
Ignited by a new computing model,
GPU Deep Learning

https://www.slideshare.net/NVIDIA/the-ai-era-ignited-by-gpu-deep-learning

Trends


100
75
50
25
1 janv. 2004
1 déc. 2008
1 nov. 2013
● Big Data
● Machine Learning
● Data Science
● Tektonik

BIG DATA LANDSCAPE 2017



INFRASTRUCTURE

HADOOP ON-PREMISE
Cloudera, Hortonworks, MapR, Pivotal, IBM InfoSphere, BlueData, Jethro

HADOOP IN THE CLOUD
Amazon Web Services, Microsoft Azure, Google Cloud Platform, IBM InfoSphere BigInsights, Treasure Data, Qubole, Altiscale, Cazena, CenturyLink

STREAMING / IN-MEMORY
Amazon Web Services, Databricks, Confluent, Striim, GridGain, Metamarkets, DataTorrent, dataArtisans, Oracle Coherence, Hazelcast, Terracotta

NOSQL DATABASES
Google Cloud Platform, Oracle, Amazon DynamoDB, Microsoft Azure, MarkLogic, mongoDB, DataStax, Aerospike, Couchbase, redislabs, influxdata

NEWSQL DATABASES
SAP HANA, Clustrix, Pivotal, NuoDB, Cockroach Labs, MemSQL, Splice Machine, MariaDB, VoltDB, citusdata, Trafodion, deepdb, paradigm4

GRAPH DBS
neo4j, IBM, Oracle, OrientDB, InfiniteGraph, Objectivity

MPP DBS
Teradata, Vertica, Netezza, Actian, Kognitio, Exasol, dremio

CLOUD EDW
Amazon Web Services, Google Cloud Platform, Microsoft Azure, Pivotal, Snowflake, Infoworks

DATA TRANSFORMATION
Talend, Pentaho, Alteryx, Trifacta, Tamr, Paxata, StreamSets, Unifi

DATA INTEGRATION
Informatica, MuleSoft, SnapLogic, Tealium, Segment, Enigma, Podium Data, Alooma, Xplenty, Zaloni, import.io, Stitch

DATA GOVERNANCE
Informatica, IBM, Skyhigh, Collibra, Alation, Waterline Data

MGMT / MONITORING
Amazon Web Services, New Relic, AppDynamics, Actifio, Datadog, Wavefront, Splunk, Unravel, Rocana, Anodot, pagerduty, Numerify

STORAGE
Amazon Web Services, Google Cloud Platform, Microsoft Azure, Alluxio, Nimble Storage, Qumulo, Coho, Panasas

CLUSTER SERVICES
Amazon Web Services, Kubernetes, Docker, Mesosphere, CoreOS, pepperdata

APP DEV
Lightbend, Keen IO, rainforest, Cask

CROWDSOURCING
Amazon Mechanical Turk, Upwork, WorkFusion, CrowdFlower

HARDWARE
Google TPU, ARM, nervana, Graphcore, Mythic, Cerebras, Nvidia, Movidius, Scortex

ANALYTICS

DATA ANALYST PLATFORMS
Microsoft, Pentaho, Alteryx, Digital Reasoning, Guavus, Ayasdi, Attivio, Datameer, Quid, ClearStory, OrigamiLogic, Interana, Bottlenose, Arimo, Endor, Mode

DATA SCIENCE PLATFORMS
IBM, KNIME, Dataiku, Domino, yhat, RapidMiner, Continuum Analytics, Alpine Data Labs, Algorithmia, Angoss

BI PLATFORMS
Microsoft, Domo, Amazon Web Services, Looker, Wave Analytics, Arcadia Data, AtScale, GoodData, Sisense

VISUALIZATION
Tableau, SAP Lumira, Google Cloud Platform, Qlik, Celonis, MapD, Periscope Data, Zepl, Chartio, plotly, Zoomdata

VERTICAL ANALYTICS
Predix, C3 IoT, Cape, Uptake, Orbital Insight, Tachyus, Alluvium, datorama

STATISTICAL COMPUTING
SAS, SPSS, Matlab

DATA SERVICES
Palantir, Mu Sigma, Opera, Data Science, kaggle, Fractal, EXL, DataKind

MACHINE LEARNING
Azure Machine Learning, Amazon Web Services, Google Cloud Platform, H2O.ai, DataRobot, context relevant, ViSenze, bonsai, DataRPM, deepsense.io, nutonian

HORIZONTAL AI
IBM Watson, Cortana, Face++, sentient, Voyager Labs, vicarious, clarifai, Affectiva, CognitiveScale, SenseTime, Numenta, Petuum, Osaro, nara, Curious AI, Blue Vision

SPEECH & NLP
Google Cloud Platform, twilio, Amazon Alexa, Narrative Science, semantic machines, Wolfram Alpha, Mobvoi 出门问问, Nuance, MindMeld, Arria, iDibon, TalkIQ, cortical.io, maluuba, snips, Gridspace, SoundHound Inc.

SEARCH
Autonomy, elastic, Oracle Endeca, Exalead, ThoughtSpot, Algolia, Lucidworks, swiftype, Maana, alphasense, Searchlink, Sinequa

LOG ANALYTICS
splunk, sumologic, loggly, kibana, logz.io

SOCIAL ANALYTICS
Hootsuite, sprinklr, Netbase, DataSift, tracx, synthesio, simplereach, bitly, predata

WEB / MOBILE / COMMERCE ANALYTICS
Google Analytics, mixpanel, Amplitude, sumall, Airtable, retention science, SigOpt, granify, custora

APPLICATIONS – ENTERPRISE

SALES
einstein, Chorus, InsideSales.com, conversica, Gong, clari, Aviso, Tact, fuse machines, Troops

MARKETING - B2B
Radius, App Annie, Everstring, Lattice, Infer, Mintigo, Reflektion, 6sense, tubular, DataFox, Engagio

MARKETING - B2C
Zeta, bloomreach, blue yonder, Persado, kahuna, ActionIQ, Bluecore, Sailthru, Quantifind, mparticle, Amplero

CUSTOMER SERVICE
Medallia, zendesk, Clarabridge, Gainsight, ClickFox, NGDATA, DigitalGenius, appuri, Automat, frame.ai, msg.ai, Intercom

HUMAN CAPITAL
HireVue, entelo, hiQ, Gigster, textio, Restless Bandit, Wade&Wendy, Clustree, Stella, pymetrics

LEGAL
Ravel, Seal, Everlaw, Judicata, Brevia, Premonition, Ross, casetext

FINANCE
anaplan, zuora, tidemark, SAP S/4HANA, Tradeshift

ENTERPRISE PRODUCTIVITY
slack, x.ai, facebook, Oracle, lumiata, diffbot, Clara, talla, butter.ai, Kasisto

BACK OFFICE AUTOMATION
HyperScience, Captricity, AppZen

SECURITY
Tanium, Cylance, StackPath, illumio, Code42, Darktrace, ThreatMetrix, DataGravity, cybereason, Vectra, CipherCloud, Guardian Analytics, Anomali, sift science, SentinelOne, SecurityScorecard, BlueTalon, Recorded Future, feedzai, Socure, Area 1 Security, Fortscale, Keybase, sqrrl, sparkcognition

APPLICATIONS – INDUSTRY

ADVERTISING
AppNexus, MediaMath, criteo, xAd, rocketfuel, Integral Ad Science, theTradeDesk, OpenX, MOAT, dstillery, Adgorithms, LiveIntent, drawbridge, gumgum, Tapad, DataXu, databerries, Appier, Dynamic Yield, Yieldmo

EDUCATION
Knewton, Clever, Declara, kidaptive, Panorama, knowre, gradescope

GOVERNMENT
Socrata, OpenGov, mark43, FiscalNote, OpenDataSoft

FINANCE - LENDING
OnDeck, Affirm, Kreditech, Avant, zestfinance, Insikt, Tala, MoneyLion, TrueAccord, trooly, cignifi, aire, Active.ai

FINANCE - INVESTING
Dataminr, Kensho, Quantopian, Numerai, iSentium, claritymoney, Algoriz, RavenPack

REAL ESTATE
Opendoor, VTS, Credifi, reonomy, Compstak

INSURANCE
Metromile, Lemonade, Cyence, Shift Technology, Tractable

HEALTHCARE
Flatiron, Kyruus, HealthTap, Metabiota, Ginger.io, Glow, Cota, zebra, ovia, AiCure, enlitic, Qventus, imagia, Arterys, BayLabs, Kang Health, Imagen, freenome

LIFE SCIENCES
23andMe, Color, iCarbonX, zymergen, BenevolentAI, Pathway Genomics, Zephyr Health, Clear Labs, Phosphorus, Citrine, twoXAR, 3Scan, Atomwise, deep genomics

TRANSPORTATION
Uber, Zoox, Tesla, nuTonomy, Clearpath, nauto, drive.ai, pilot.ai, Optimus Ride, Otto, nexar, Civil Maps, comma.ai, NIO, netradyne

AGRICULTURE
Farmers Business Network, FarmersEdge, FarmLogs, Blue River, mavrx, TerrAvion, prospera

COMMERCE
instacart, Stitch Fix, RetailNext, HowGood

OTHER
eHarmony, stem, rethink robotics, Amper Music, Sight Machine, hopper, Boxever, celect, Verdigris, duetto, Unbabel, Jukedeck, Second Spectrum

CROSS-INFRASTRUCTURE/ANALYTICS
Amazon Web Services, Google Cloud Platform, Microsoft, IBM, SAP, Hewlett Packard Enterprise, SAS, 1010data, vmware, TIBCO, Teradata, Oracle, NetApp

OPEN SOURCE

FRAMEWORK
Hadoop HDFS, Hadoop MapReduce, Flink, YARN, Tez, Spark, Mesos, CDAP

QUERY / DATA FLOW
Spark SQL, presto, Hive, SlamData, Apache Drill, Google Cloud Dataflow

DATA ACCESS
cassandra, nifi, mongoDB, SciDB, CouchDB, OpenTSDB, riak, Apache HBase, Spanner, Accumulo

COORDINATION
talend, Apache Zookeeper, Apache Ambari

STREAMING
Spark, Apex, Flink, Beam, kafka, druid, Storm

STAT TOOLS
python, R, ScalaLab, NumPy, SciPy

AI / MACHINE LEARNING / DEEP LEARNING
TensorFlow, theano, torch, MADlib, learn, Caffe, CNTK, DMTK, Keras, PaddlePaddle, Apache SINGA, OpenAI, mxnet, Velles, Weka, neon, FeatureFu, Chainer, DIMSUM, DSSTNE, mllib, DL4J, Mahout, Aerosolve

SEARCH
elasticsearch, Solr, Lucene

LOG ANALYSIS
elasticsearch, kibana, logstash

VISUALIZATION
Beaker, Rodeo

COLLABORATION
jupyter, Zeppelin, Anaconda

SECURITY
Apache Ranger, Knox, Sentry

DATA SOURCES & APIS

HEALTH
Apple, Jawbone, Validic, practice fusion, fitbit, Garmin, Human API, kinsa

IOT
GE Digital, Uptake, ThingWorx, helium, samsara, Augury, estimote

FINANCIAL & ECONOMIC DATA
Bloomberg, Thomson Reuters, Dow Jones, S&P Capital IQ, CB Insights, xignite, quandl, Yodlee, Premise, estimize, Second Measure, Eagle Alpha, StockTwits, Plaid, mattermark, Thinknum

AIR / SPACE / SEA
Planet Labs, Airware, Skycatch, spire, Airbotics, Descartes Labs, Understory, Windward, TellusLabs, DroneDeploy, MarineTraffic

PEOPLE / ENTITIES
acxiom, Experian, Epsilon, InsideView, Crimson Hexagon, Basis, quantcast, SafeGraph

LOCATION INTELLIGENCE
Foursquare, MapAnything, Sense360, PlaceIQ, esri, factual, Carto, Mapillary, Streetline

OTHER
qualtrics, Data.gov, data.world, panjiva, enigma

DATA RESOURCES

INCUBATORS & SCHOOLS
Pluralsight, General Assembly, galvanize, DataCamp, DataElite, Insight, The Data Incubator, Metis

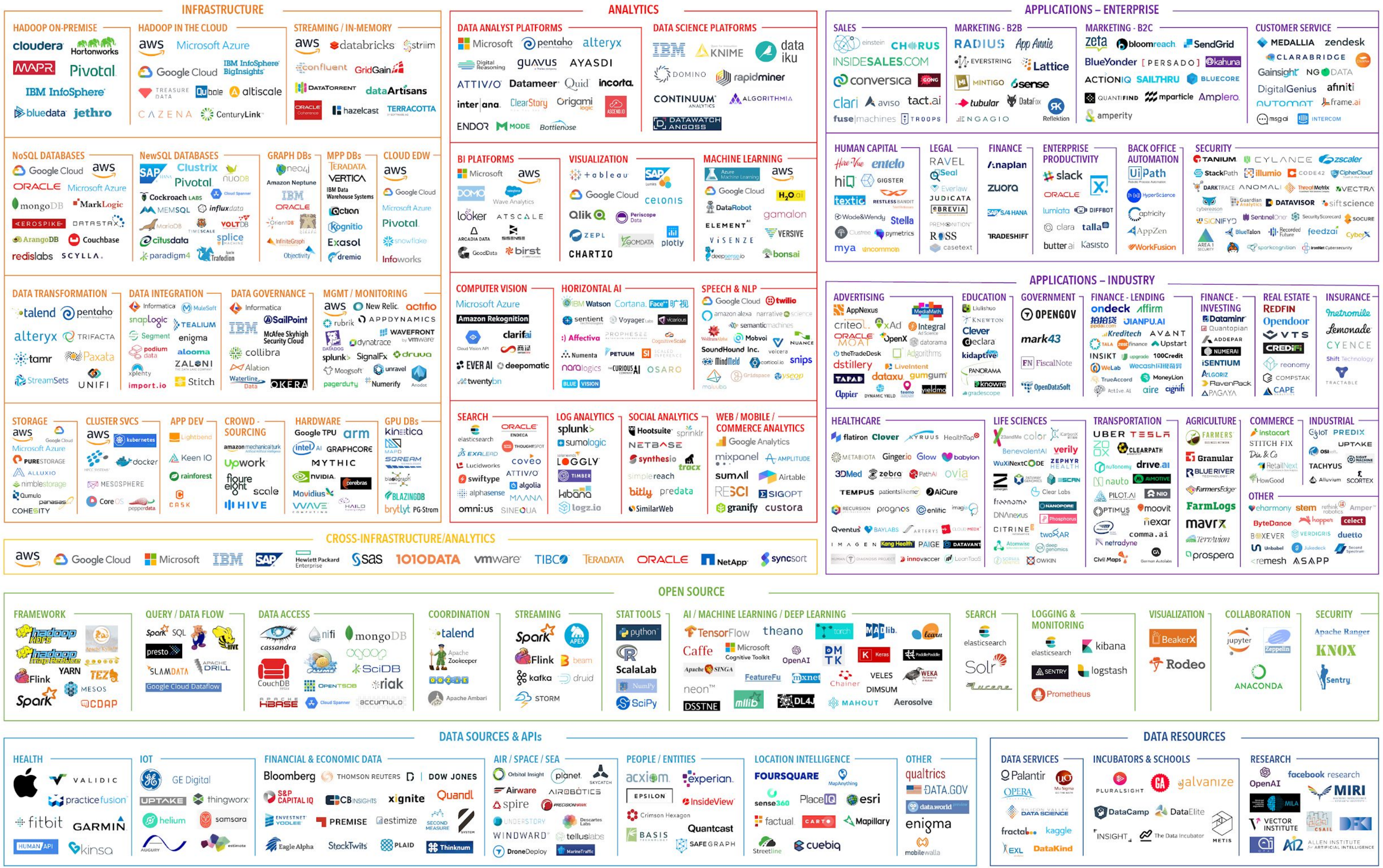
RESEARCH
facebook research, OpenAI, MIRI, CSAIL, DFKI, AI2 Allen Institute for Artificial Intelligence

V2 – Last updated 5/3/2017

© Matt Turck (@mattturck), Jim Hao (@jimrhao), & FirstMark (@firstmarkcap) mattturck.com/bigdata2017

FIRSTMARK EARLY STAGE VENTURE CAPITAL

BIG DATA & AI LANDSCAPE 2018



INFRASTRUCTURE

HADOOP ON-PREMISE: cloudera, Hortonworks, MAPR, Pivotal, IBM InfoSphere, bluedata, jethro

HADOOP IN THE CLOUD: aws, Microsoft Azure, Google Cloud, IBM InfoSphere BigInsights, Treasure Data, Qubole, altiscale, CAZENA, CenturyLink

STREAMING / IN-MEMORY: aws, databricks, striim, confluent, GridGain, DataTorrent, dataArtisans, Oracle Coherence, hazelcast, TERRACOTTA

NoSQL DATABASES: Google Cloud, aws, ORACLE, Microsoft Azure, mongoDB, MarkLogic, AEROSPIKE, DATASTAX, ArangoDB, Couchbase, redislabs, SCYLLA

NewSQL DATABASES: SAP HANA, Clustrix, Pivotal, NuoDB, Cockroach Labs, Cloud Spanner, MEMSQL, influxdata, MariaDB, TimeScale, VoltDB, citusdata, splice machine, paradigm4, Trafodion

GRAPH DBs: neo4j, Amazon Neptune, IBM, ORACLE, OrientDB, InfiniteGraph, Objectivity

MPP DBs: TERADATA, VERTICA, IBM Data Warehouse Systems, Actian, Kognitio, Exasol, dremio

CLOUD EDW: aws, Google Cloud, Microsoft Azure, Pivotal, snowflake, Infoworks

DATA TRANSFORMATION: talend, pentaho, alteryx, TRIFACTA, tamr, Paxata, StreamSets, UNIFI

DATA INTEGRATION: Informatica, MuleSoft, snapLogic, TEALIUM, Segment, enigma, podium data, alooma, xplenty, ZALONI, import.io, Stitch

DATA GOVERNANCE: Informatica, IBM, SailPoint, McAfee Skyhigh Security Cloud, collibra, Alation, Waterline Data, OKERA

MGMT / MONITORING: aws, New Relic, actifio, rubrik, APPDYNAMICS, DATADOG, WAVEFRONT by vmware, dynatrace, splunk, SignalFx, druva, Moogsoft, unravel, pagerduty, Numerify, Anodot

STORAGE: aws, Google Cloud, Microsoft Azure, PURESTORAGE, ALLUXIO, nimblestorage, Qumulo, panasas, COHESITY

CLUSTER SVCS: aws, kubernetes, HPE, docker, MESOSPHERE, CoreOS, pepperdata

APP DEV: Lightbend, Keen IO, rainforest, CASK

CROWD-SOURCING: amazon mechanical turk, upwork, MYTHIC, figure eight, scale, HIVE

HARDWARE: Google TPU, arm, intel AI, GRAPHCORE, NVIDIA, cerebras, Movidius, WAVE COMPUTING, HALO

GPU DBs: kinetica, MAPD, SQREAM, blazegraph, BLAZINGDB, brytlyt, PG-Strom

ANALYTICS

DATA ANALYST PLATFORMS: Microsoft, pentaho, alteryx, Digital Reasoning, guavus, AYASDI, ATTIVIO, Datameer, Quid, incorta, inter|ana, ClearStory, Origami Logic, ASCEND.IO, ENDOR, MODE, Bottlenose

DATA SCIENCE PLATFORMS: IBM, KNIME, dataiku, DOMINO, rapidminer, CONTINUUM ANALYTICS, ALGORITHMIA, DATAWATCH ANGOSS

BI PLATFORMS: Microsoft, aws, DOMO, salesforce Wave Analytics, looker, ATSCALE, ARCADIA DATA, SISENSE, GoodData, birst

VISUALIZATION: tableau, SAP Lumira, Google Cloud, celonis, Qlik, Periscope Data, ZEPL, ZOOMDATA, plotly, CHARTIO

MACHINE LEARNING: Azure Machine Learning, aws, Google Cloud, H2O.ai, DataRobot, gamalon, ELEMENT AI, VERSIVE, VISENZE, deepsense.io, bonsai

COMPUTER VISION: Microsoft Azure, Amazon Rekognition, clarifai, Cloud Vision API, 商汤, EVER AI, deepomatic, twentybn

HORIZONTAL AI: IBM Watson, Cortana, Face++, 旷视, sentient, Voyager Labs, vicarious, Affectiva, PROPHESEE, CognitiveScale, Numenta, PETUUM, SparkCognition, naralogics, CURIOUS AI, OSARO, BLUE VISION

SPEECH & NLP: Google Cloud, twilio, amazon alexa, narrative science, semantic machines, SoundHound Inc., Mobvoi, NUANCE, MindMeld, voicera, snips, Gridspace, ysg

SEARCH: ORACLE ENDECA, elasticsearch, THOUGHTSPOT, EXALEAD, Lucidworks, coveo, swiftype, ATTIVIO, alphasense, algolia, omni:us, MAANA, SINEQUA

LOG ANALYTICS: splunk, sumologic, LOGGLY, TIMBER, kibana, logz.io

SOCIAL ANALYTICS: Hootsuite, sprinklr, NETBASE, synthesio, tracx, simplereach, bitly, predata, SimilarWeb

WEB / MOBILE / COMMERCE ANALYTICS: Google Analytics, mixpanel, AMPLITUDE, sumAll, Airtable, RESCI, SIGOPT, granify, custora

APPLICATIONS – ENTERPRISE

SALES: einstein, CHORUS, INSIDESALES.COM, conversica, GONG, clari, aviso, tact.ai, fuse machines, TROOPS

MARKETING - B2B: RADIUS, App Annie, EVERSTRING, Lattice, MINTIGO, 6sense, tubular, DataFox, Reflektion, ENGAGIO

MARKETING - B2C: zeta, bloomreach, SendGrid, BlueYonder, PERSADO, kahuna, ACTIONIQ, SAILTHRU, BLUECORE, QUANTIFIND, mparticle, Amplero, amperity

CUSTOMER SERVICE: MEDALLIA, zendesk, CLARABRIDGE, Gainsight, NG DATA, DigitalGenius, afiniti, AUTOMAT, frame.ai, msg.ai, INTERCOM

HUMAN CAPITAL: HireVue, entelo, hiQ, GIGSTER, textio, RESTLESS BANDIT, Wade&Wendy, Stella, Custora, pymetrics, mya, uncommon

LEGAL: RAVEL, Seal, Everlaw, JUDICATA, BREVIA, PREMONITION, ROSS, casetext

FINANCE: Anaplan, zuora, SAP S/4 HANA, TRADESHIFT

ENTERPRISE PRODUCTIVITY: slack, X.ai, ORACLE, lumiata, DIFFBOT, clara, talla, butter.ai, Kasisto

BACK OFFICE AUTOMATION: UiPath, HyperScience, apptricity, AppZen, WorkFusion

SECURITY: TANIUM, CYLANCE, zscaler, StackPath, illumio, CODE42, CipherCloud, DARKTRACE, ANOMALI, ThreatMetrix, VECTRA, cybereason, Guardian Analytics, DATAVISOR, siftscience, SIGNIFYD, SentinelOne, SecurityScorecard, SOCURE, AREA 1 SECURITY, BlueTalon, Recorded Future, feedzai, CyberX, sparkcognition, IronNet Cybersecurity

APPLICATIONS – INDUSTRY

ADVERTISING: AppNexus, MediaMath, criteo, RocketFuel, Integral Ad Science, ORACLE MOAT, Openx, datorama, theTradeDesk, Algorithmia, dstillery, Liveintent, TAPAD, dataxu, gumgum, Appier, DYNAMIC YIELD, teemo, yieldmo

EDUCATION: Liulishuo, KNEWTON, Clever, declara, kidaptive, PANORAMA, knowre, gradescope

GOVERNMENT: OPENGOV, mark43, FiscalNote, OpenDataSoft

FINANCE - LENDING: ondeck, affirm, 拍拍贷, JIANPU.AI, Kreditech, AVANT, TALA, ZestFinance, Upstart, INSIKT, upgrade, 100Credit, WeLab, Wecash闪银, TrueAccord, MoneyLion, Active.Ai, aire, cignifi

FINANCE - INVESTING: Dataminr, Quantopian, ADDEPAR, NUMERAI, iSENTIUM, ALGORIZ, RavenPack, APAGAYA

REAL ESTATE: REDFIN, Opendoor, VTS, CREDIFI, reonomy, COMPSTAK, CAPE

INSURANCE: metromile, Lemonade, CYENCE, Shift Technology, TRACTABLE

HEALTHCARE: flatiron, Clover, KYRUUS, HealthTap, METABIOTA, Ginger.io, Glow, babylon, 3DMed, zebra, PathAI, ovia, TEMPUS, patientslikeme, AiCure, RECURSION, prognos, enlitic, imagia, Qventus, BAYLABS, ARTERYS, Human Dx, Kang Health, PAIGE, DATAVANT, innovaccer, LeanTaaS

LIFE SCIENCES: 23andMe, color, icarbonx, BenevolentAI, verily, WuXiNextCODE, ZEPHYR HEALTH, Zymergen, atomwise, 3SCAN, Clear Labs, freenome, NANOPORE, DNAnexus, Phosphorus, CITRINE, twoXAR, Atomwise, deep genomics, OWKIN

TRANSPORTATION: UBER, TESLA, ZOOX, CLEARPATH, nuTonomy, drive.ai, nauto, AMOTIVE, PILOT.AI, NIO, OPTIMUS RIDE, moovit, nexar, comma.ai, netradyne, Civil Maps, German Autolabs

AGRICULTURE: FARMERS BUSINESS NETWORK, Granular, BLUE RIVER TECHNOLOGY, FarmersEdge, FarmLogs, mavrx, TerrAvion, prospera

COMMERCE: instacart, STITCH FIX, Dia & Co, RetailNext, HowGood

INDUSTRIAL: GE IoT, PREDIX, UPTAKE, osisoft, SIGHT MACHINE, TACHYUS, Alluvium, SCORTEX

OTHER: eharmony, stem, rethink robotics, Amper, ByteDance, hopper, celect, BOXEVER, VERDIGRIS, duetto, Unbabel, Jukedeck, Second Spectrum, remesh, ASAPP

CROSS-INFRASTRUCTURE/ANALYTICS

aws, Google Cloud, Microsoft, IBM, SAP, Hewlett Packard Enterprise, sas, 1010DATA, vmware, TIBCO, TERADATA, ORACLE, NetApp, syncsort

OPEN SOURCE

FRAMEWORK: hadoop HDFS, Apache Kylin, hadoop MapReduce, YARN, TEZ, Flink, MESOS, Spark, CDAP

QUERY / DATA FLOW: Spark SQL, Pig, HIVE, presto, SLAMDATA, APACHE DRILL, Google Cloud Dataflow

DATA ACCESS: cassandra, nifi, mongoDB, ooooo, SciDB, CouchDB, OPENTSDB, riak, HBASE, Cloud Spanner, accumulo

COORDINATION: talend, Apache Zookeeper, OOZIE, Apache Ambari

STREAMING: Spark, APEX, Flink, beam, kafka, druid, STORM

STAT TOOLS: python, R, ScalaLab, NumPy, SciPy

AI / MACHINE LEARNING / DEEP LEARNING: TensorFlow, theano, torch, MADlib, Caffe, Microsoft Cognitive Toolkit, OpenAI, DMTK, Keras, PaddlePaddle, learn, Apache SINGA, FeatureFu, mxnet, Chainer, VELES, WEKA, neon, DSSTNE, mllib, DL4J, MAHOUT, DIMSUM, Aerosolve

SEARCH: elasticsearch, Solr, Lucene

LOGGING & MONITORING: elasticsearch, kibana, SENTRY, logstash, Prometheus

VISUALIZATION: BeakerX, Rodeo

COLLABORATION: jupyter, Zeppelin, ANACONDA

SECURITY: Apache Ranger, KNOX, Sentry

DATA SOURCES & APIs

HEALTH: Apple, VALIDIC, practice fusion, fitbit, GARMIN, HUMAN API, kinsa

IOT: GE Digital, UPTAKE, thingworx, helium, samsara, AUGURY, estimote

FINANCIAL & ECONOMIC DATA: Bloomberg, THOMSON REUTERS, DOW JONES, S&P CAPITAL IQ, CB INSIGHTS, xignite, Quandl, INVESTNET YODLEE, PREMISE, estimize, SECOND MEASURE, Eagle Alpha, StockTwits, PLAID, Thinknum

AIR / SPACE / SEA: Orbital Insight, planet, SKYCATCH, Airware, AIRBOTICS, spire, PRECISIONHAWK, UNDERSTORY, Descartes Labs, WINDWARD, tellus labs, DroneDeploy, MarineTraffic

PEOPLE / ENTITIES: acxiom, experian, EPSILON, InsideView, Crimson Hexagon, BASIS, Quantcast, SAFEGRAPH

LOCATION INTELLIGENCE: FOURSQUARE, MapAnything, sense360, Placed, esri, factual, CARTO, Mapillary, Streetline, cuebiq

OTHER: qualtrics, DATA.GOV, data.world, enigma, mobilewalla

DATA RESOURCES

DATA SERVICES: Palantir, Mu Sigma, OPERA, DATA SCIENCE, fractal, kaggle, EXL, DataKind

INCUBATORS & SCHOOLS: PLURALSIGHT, GA, galvanize, DataCamp, DataElite, INSIGHT, The Data Incubator, METIS

RESEARCH: facebook research, OpenAI, MIRI, MILA, VECTOR INSTITUTE, CSAIL, DFKI, AI2, ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

V1 – Last updated 6/19/2018

© Matt Turck (@mattturck), Demilade Obayomi (@demi_obayomi), & FirstMark (@firstmarkcap) mattturck.com/bigdata2018

FIRSTMARK EARLY STAGE VENTURE CAPITAL

Data Scientist: The Sexiest Job of the 21st Century

EVERYTHING STARTS WITH A WHY
Y
ALEXANDRE ROBICQUET
ARTIFICIAL INTELLIGENCE RESEARCHER
Y, THE NEW FRAGRANCE FOR MEN
YVESSAINTLAURENT
YSL
yslbeauty.com

https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

Keep Calm with Michael Jordan

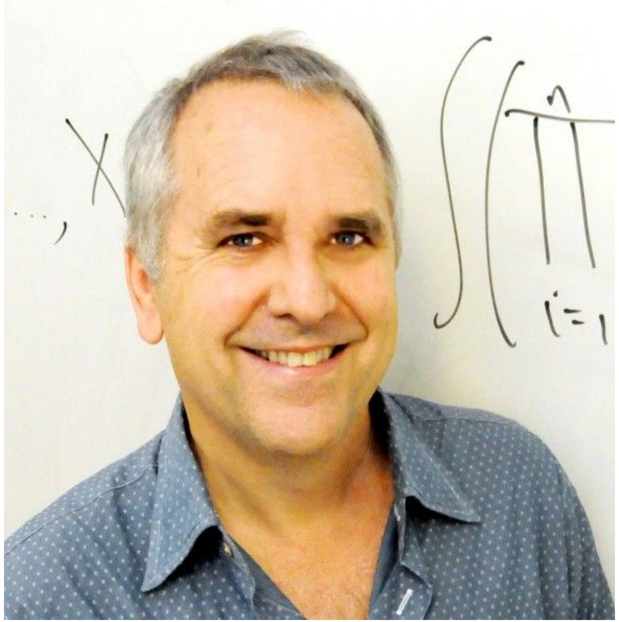
Artificial Intelligence—The Revolution Hasn’t Happened Yet

https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7

Keep Calm with Michael Jordan

Artificial Intelligence — The Revolution Hasn’t Happened Yet

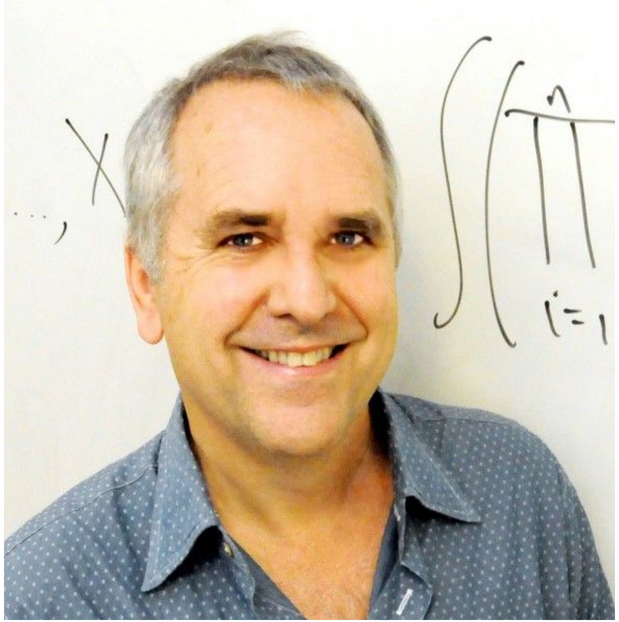
https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7



Keep Calm with Michael Jordan

Artificial Intelligence—The Revolution Hasn't Happened Yet

- We are currently building Machine Learning blocks
- Blocks are being put together in ad-hoc ways
- We are far from what we call "AI"



https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7

Machine Learning Craftsmanship

Kelvin MOUTET

Machine Learning Engineer @ Prevision.io

Software Craftsmanship

As aspiring Software Craftsmen we are raising the bar of professional software development by practicing it and helping others learn the craft. Through this work we have come to value:

Not only working software ,
but also well-crafted software
Not only responding to change,
but also steadily adding value
Not only individuals and interactions,
but also a community of professionals
Not only customer collaboration,
but also productive partnerships

That is, in pursuit of the items on the left we have found the items on the right to be indispensable.

http://manifesto.softwarecraftsmanship.org/

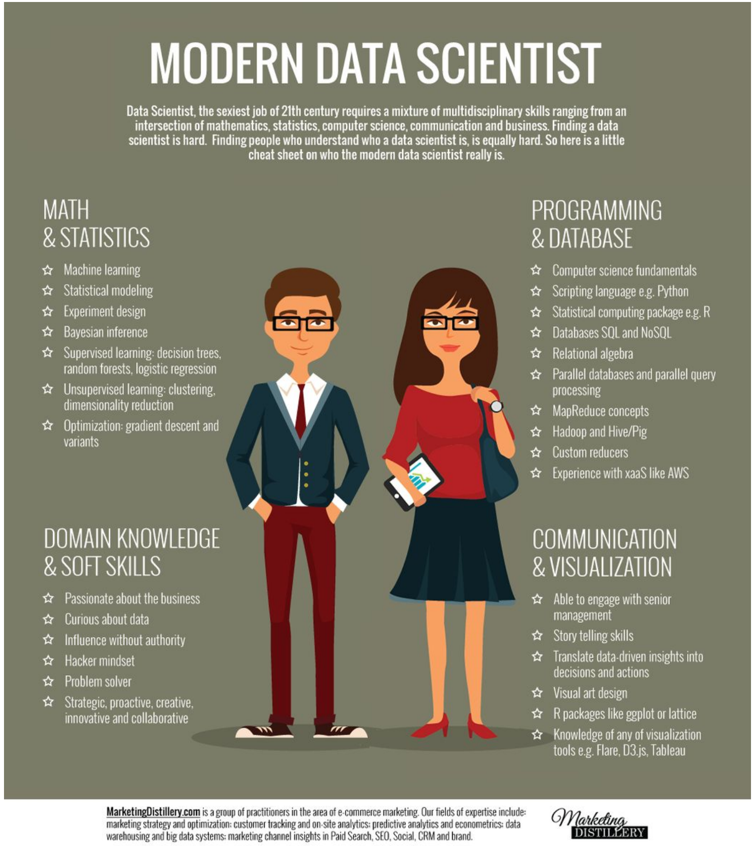
PLAN

2 parts only !

1) Part 1

Machine Learning Bare Necessities

- Data Scientists !



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

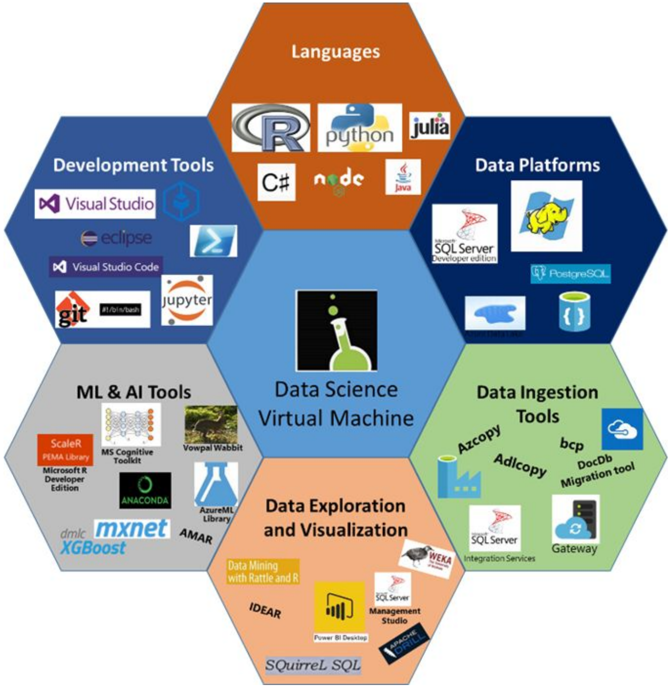
- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization: customer tracking and on-site analytics: predictive analytics and econometrics: data warehousing and big data systems: marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing DISTILLERY

Machine Learning Bare Necessities

- Data Scientists
- Tools !



http://blog.revolutionanalytics.com/2017/03/dsvm-linux-updates.html

Machine Learning Bare Necessities

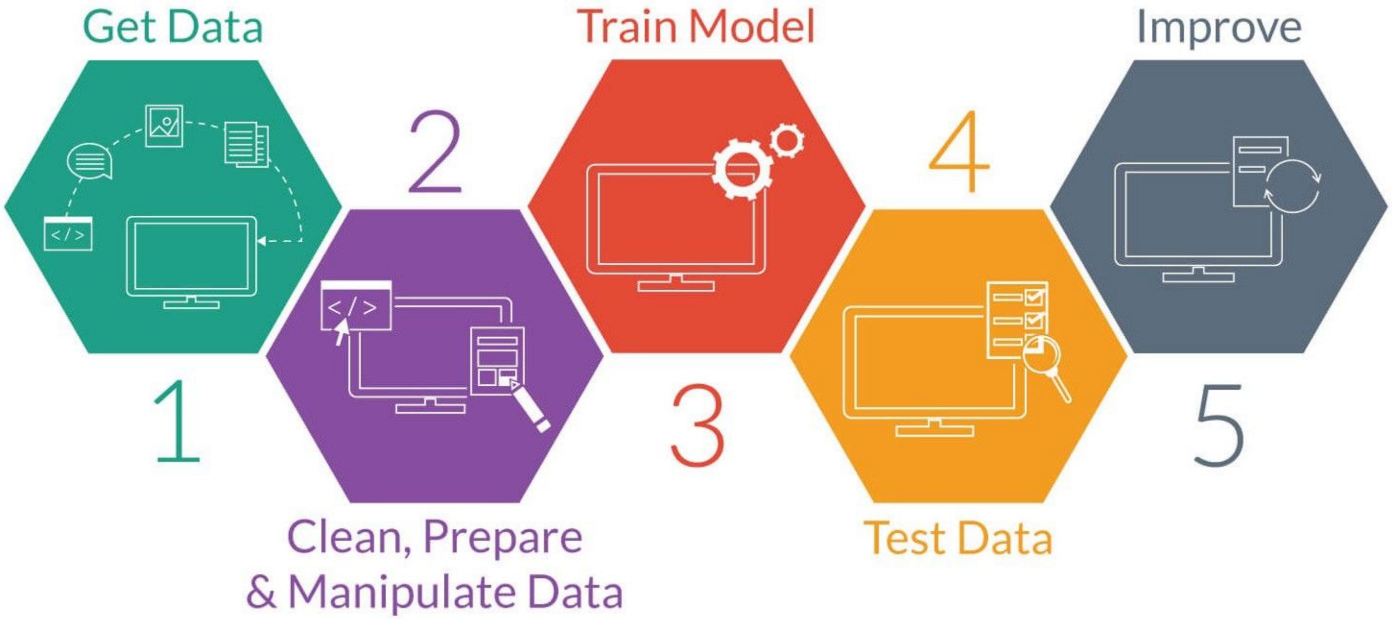
- Data Scientists
- Tools
- Data !

Machine Learning Bare Necessities

- Data Scientists
- Tools
- Data
- BIG DATA !

Let’s go !

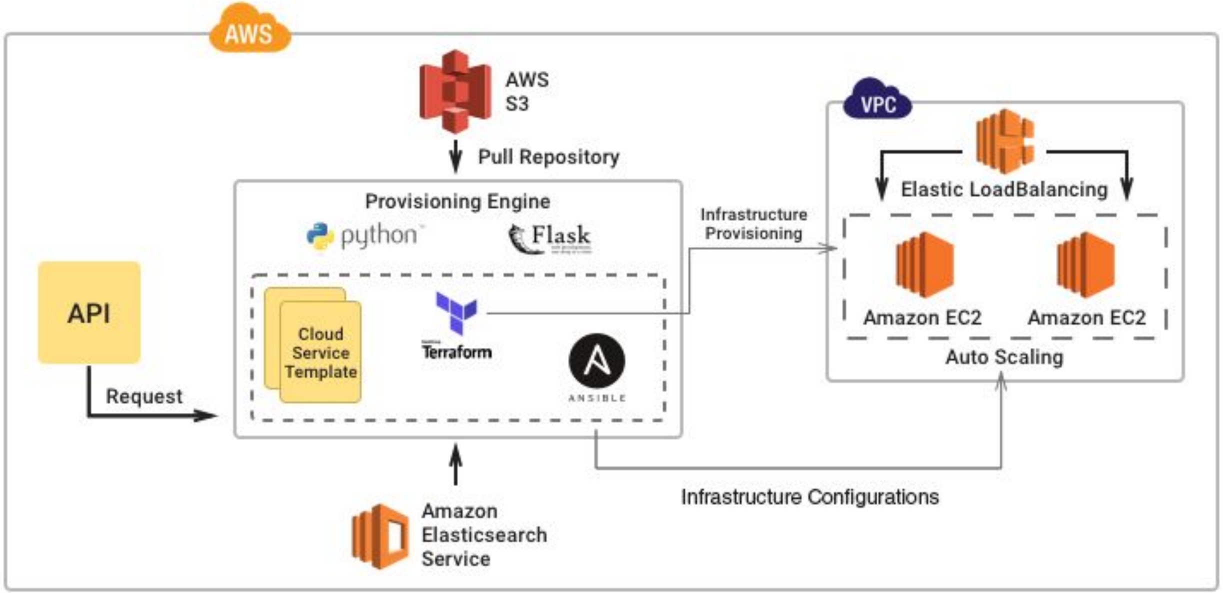
Machine Learning core steps



https://upxacademy.com/introduction-machine-learning/

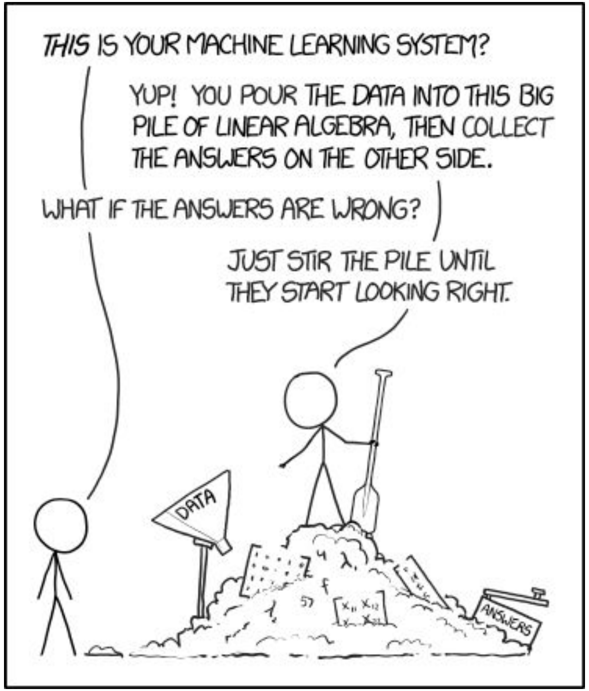
Go to Production !

Machine Learning Infrastructure

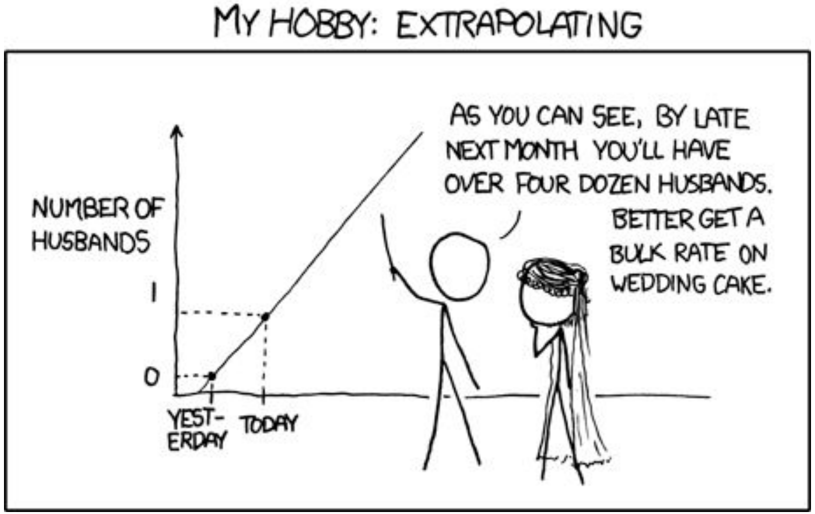


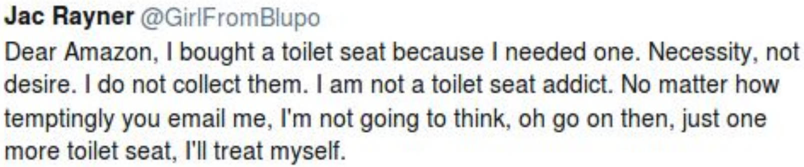
https://www.opcito.com/build-infrastructure-operations-using-aws-services-terraform-ansible-data-analytics-company

Welcome to AI ERA !!!


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Machine Learning “Success” Stories

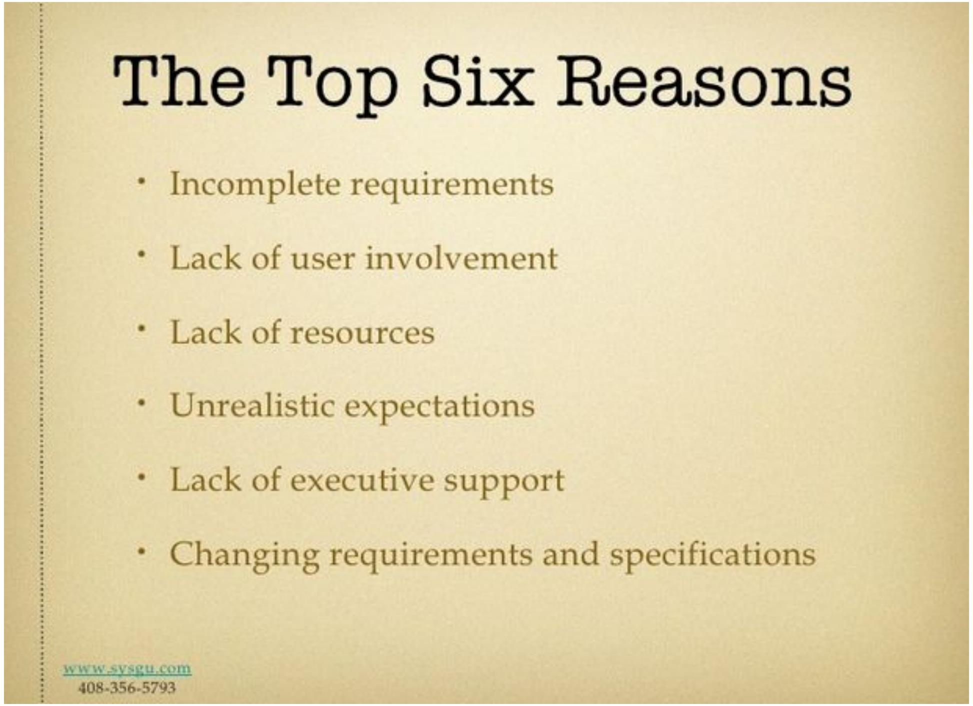

MY HOBBY: EXTRAPOLATING
NUMBER OF HUSBANDS
1
0
YESTERDAY
TODAY
AS YOU CAN SEE, BY LATE NEXT MONTH YOU'LL HAVE OVER FOUR DOZEN HUSBANDS. BETTER GET A BULK RATE ON WEDDING CAKE.


Jac Rayner @GirlFromBlupo

Dear Amazon, I bought a toilet seat because I needed one. Necessity, not desire. I do not collect them. I am not a toilet seat addict. No matter how temptingly you email me, I'm not going to think, oh go on then, just one more toilet seat, I'll treat myself.

Why does it fail ?

Software project failure !



The Top Six Reasons

- Incomplete requirements
- Lack of user involvement
- Lack of resources
- Unrealistic expectations
- Lack of executive support
- Changing requirements and specifications

www.sysgu.com
408-356-5793

https://www.slideshare.net/bstaud/six-reasons-for-software-project-failure-presentation

Machine Learning Project Failure (not exhaustive)

- Problem definition is bad or unclear (incomplete requirements)
- Scope is too large (unrealistic expectations)
- Overconfidence on data scientists (lack of resources)
- (Big) Data is too small or bad ! (lack of resources)
- True AI is not here (unrealistic expectations)
- Pessimistic on core business knowledge (lack of user involvement)
- Tools/ML Driven Development instead of Core Business Development
- Machine Learning algorithms understanding/coverage

1) ~~Part 1~~
2) Part 2

2) Machine Learning Craftsmanship

Software Craftsmanship

As aspiring Software Craftsmen we are raising the bar of professional software development by practicing it and helping others learn the craft. Through this work we have come to value:

Not only working software ,
but also well-crafted software
Not only responding to change,
but also steadily adding value
Not only individuals and interactions,
but also a community of professionals
Not only customer collaboration,
but also productive partnerships

That is, in pursuit of the items on the left we have found the items on the right to be indispensable.

http://manifesto.softwarecraftsmanship.org/

Machine Learning Craftsmanship

As aspiring Software Craftsmen we are raising the bar of professional software development by practicing it and helping others learn the craft. Through this work we have come to value:

Not only working software ,

but also well-crafted software

Not only responding to change,

but also steadily adding value

Not only individuals and interactions,

but also a community of professionals

Not only customer collaboration,

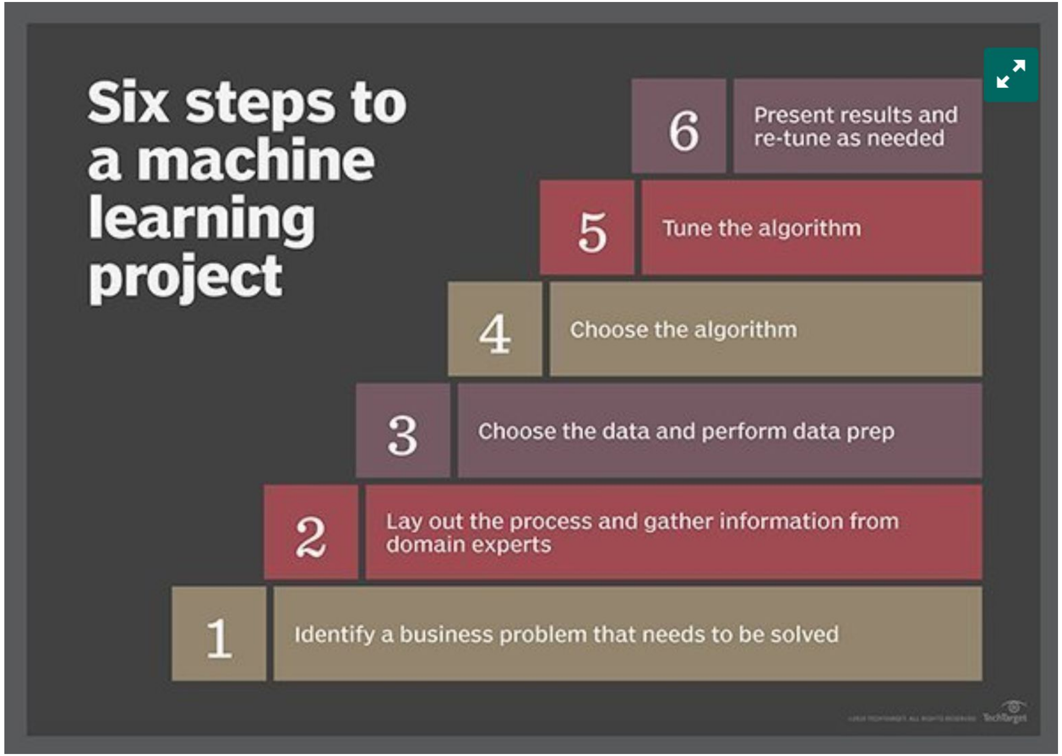
but also productive partnerships

That is, in pursuit of the items on the left we have found the items on the right to be indispensable.

Machine Learning Craftsmanship

Mindset !

A good way to start a Machine Learning project



Six steps to a machine learning project

6 Present results and re-tune as needed

5 Tune the algorithm

4 Choose the algorithm

3 Choose the data and perform data prep

2 Lay out the process and gather information from domain experts

1 Identify a business problem that needs to be solved

https://searchenterpriseai.techtarget.com/tip/How-to-use-machine-learning-to-build-a-predictive-algorithm

Examples !

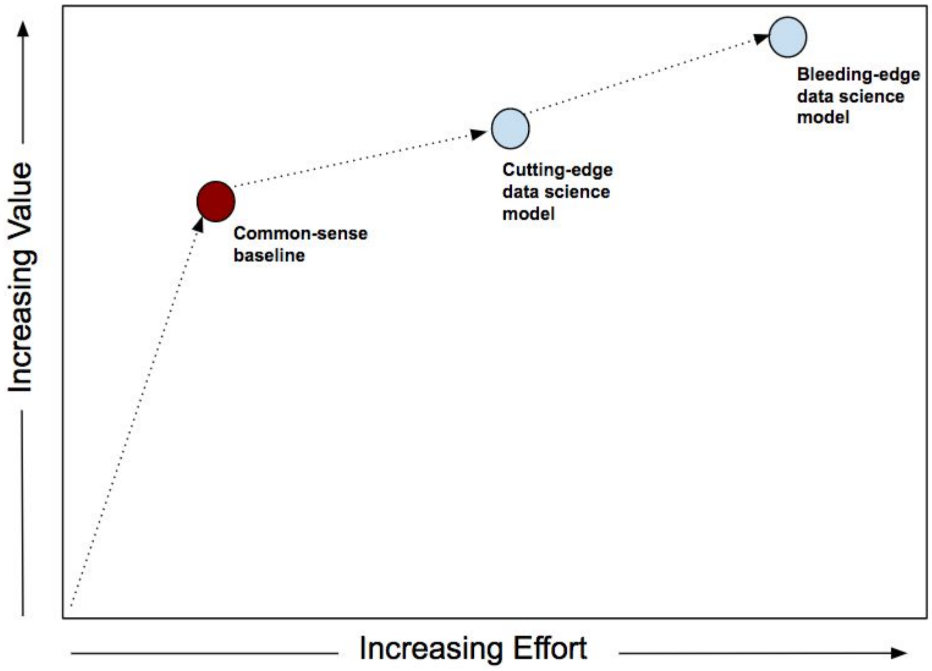
Machine Learning problem solving

Inspired by those two articles :

https://towardsdatascience.com/1-year-doing-data-science-in-the-real-world-54f49b591991 by **Jonny Brooks-Bartlett**, Data scientist Deliveroo

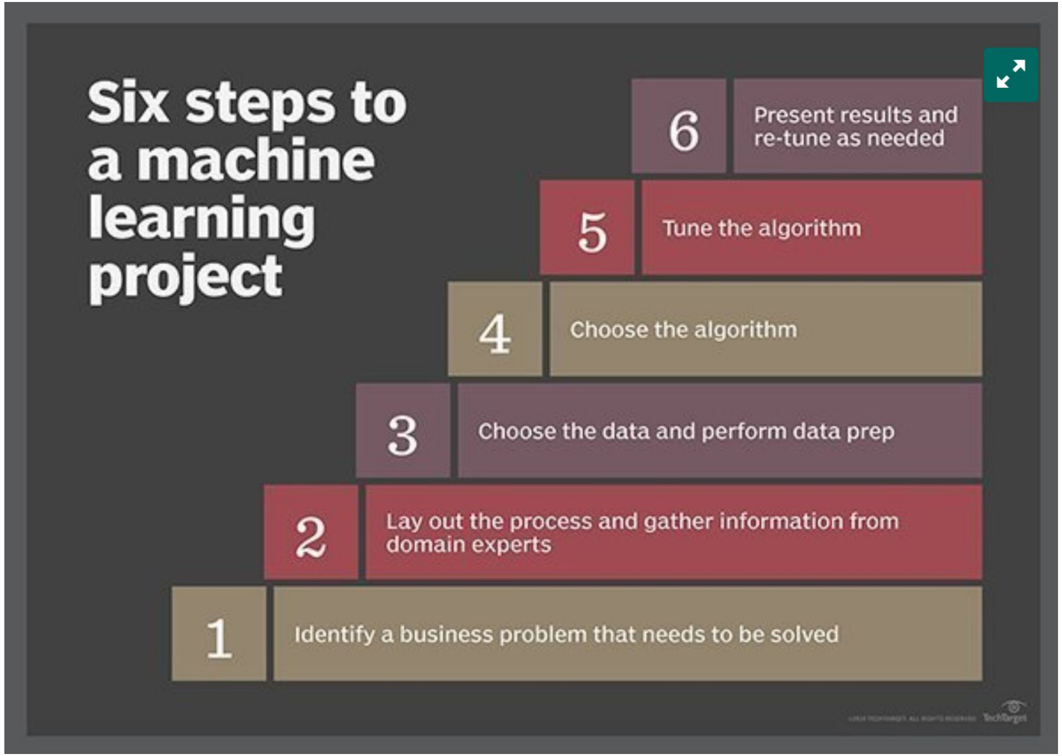
https://towardsdatascience.com/first-create-a-common-sense-baseline-e66dbf8a8a47 by **Rama Ramakrishnan**, Senior Vice President, Salesforce.

Identify your problem then build value !


Bleeding-edge data science model
Cutting-edge data science model
Common-sense baseline
Increasing Value
Increasing Effort

https://towardsdatascience.com/first-create-a-common-sense-baseline-e66dbf8a8a47

A good way to start a ML project



Six steps to a machine learning project

1 Identify a business problem that needs to be solved

2 Lay out the process and gather information from domain experts

3 Choose the data and perform data prep

4 Choose the algorithm

5 Tune the algorithm

6 Present results and re-tune as needed

https://searchenterpriseai.techtarget.com/tip/How-to-use-machine-learning-to-build-a-predictive-algorithm

Customers targeting

Send an advertising of your latest new products, at most, to 100,000 customers from the database.

How do you solve this ?

Customers targeting

Pick the most **loyal** customers

a) Shop a lot (frequency) and b) spends (money) a lot over a period (year, month)

c) Recently shopped with you (number of day, week since last transaction)

Decile the customer databases, sort and pick the 100 000 first !

This is a core business tools called **Recency-Frequency-Monetary (RFM) heuristic**

Product recommendations

Build a personalized product recommendation zone that will be displayed on the home page.

How do you do it ?

Product recommendations

Transform the problem to : What’s the simplest recommandation you can show visitors?

- Top selling products over all or by categories on a time window
- It works for all visitors (with or without informations)
- Personalisation with tweak (remember categories visited)

This is a good baseline before doing ML based methods

What next ?

Machine Learning common issues !

Machine Learning algorithms issues

[Data] Cold Start / Warm Start

[Data] Conceptual Drift / Covariate shift

[Model] No free lunch Theorem against you

[Data/Model] Reproduce

[Data/Model] Data Bias / Leak of information

[Data/Model] Cross validation and test validation

Machine Learning project issues

[Model Selection] Usability (under core business constraints)
[Model selection] Core business metrics vs ML metrics !
[Model evaluation] A/B testing

[Production] Automatise update mode
[Production] Monitoring (speed, failure, user feedback)

Conclusion

❏ Talk with core business, as software engineers do, and build a common sense baseline !

❏ Fear over engineering and useless Machine Learning

❏ Keep learning :
  - Mooc, tutorials, books, articles, talks
  - Forum, blogs, twitter, people

Machine Learning Craftsmanship

Kelvin MOUTET

Machine Learning Engineer @ Prevision.io

iAdvize

dictanova

CogniTalk

EPITECH.

L'ECOLE DE L'INNOVATION ET DE L'EXPERTISE INFORMATIQUE

ML WEEK

THE MACHINE LEARNING WEEK