



2017

Global Azure **BOOTCAMP**

@Microsoft

April 22 | Athens, GR

Data Science with Azure Machine Learning and R

Christos Charmatzis

Data Scientist

R&D, MarineTraffic

xr_xarmatzis@outlok.com @xarmatzis

Objectives

Introduce Microsoft
data science

Focus on Azure ML + R +
Jupyter Notebook

Why?

Discover **reason** behind success, failure

Understand customers, products, yourself

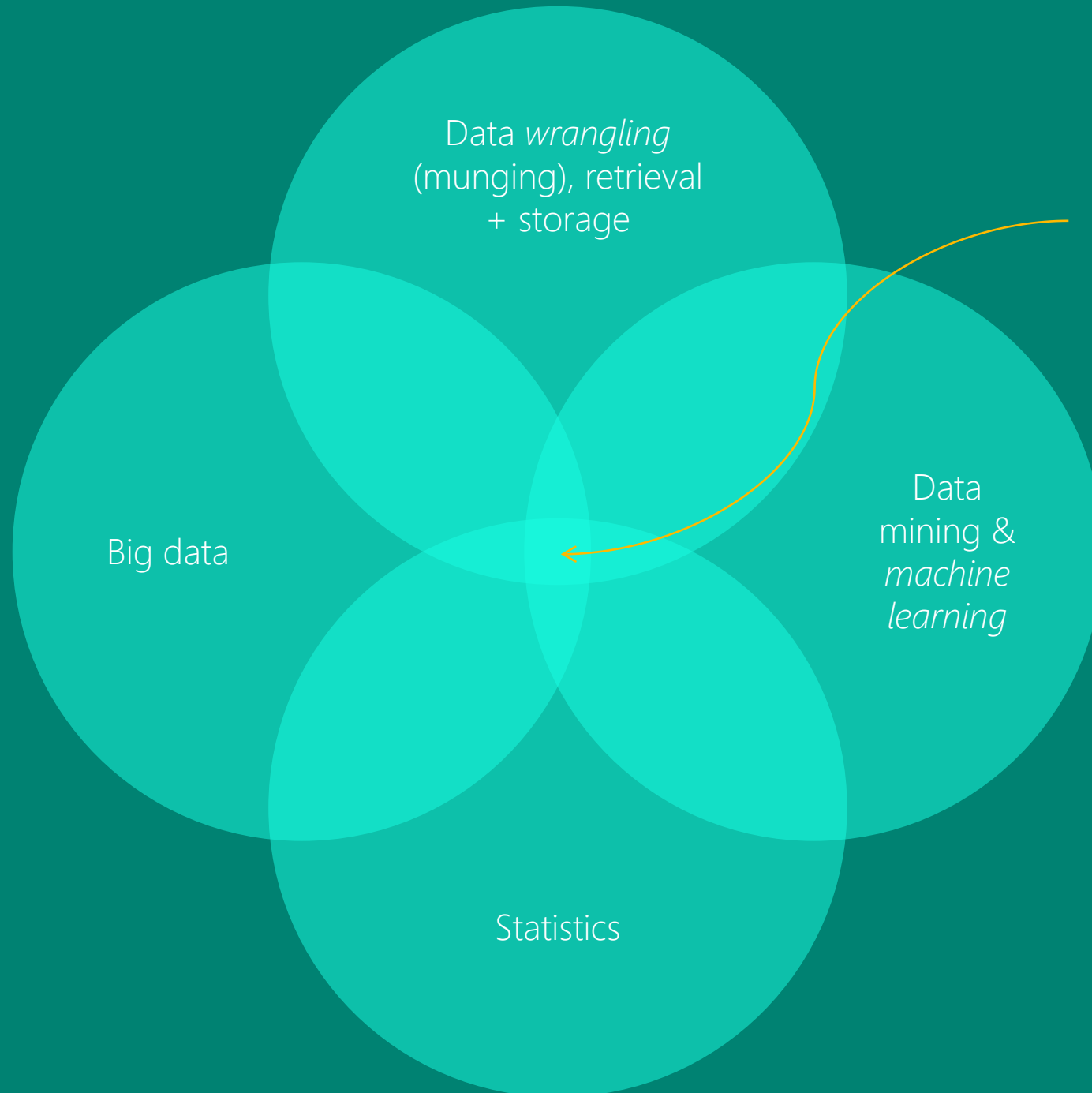
Plan future

Experiment meaningfully

Improve **performance**

Run on **analytics**

Data science

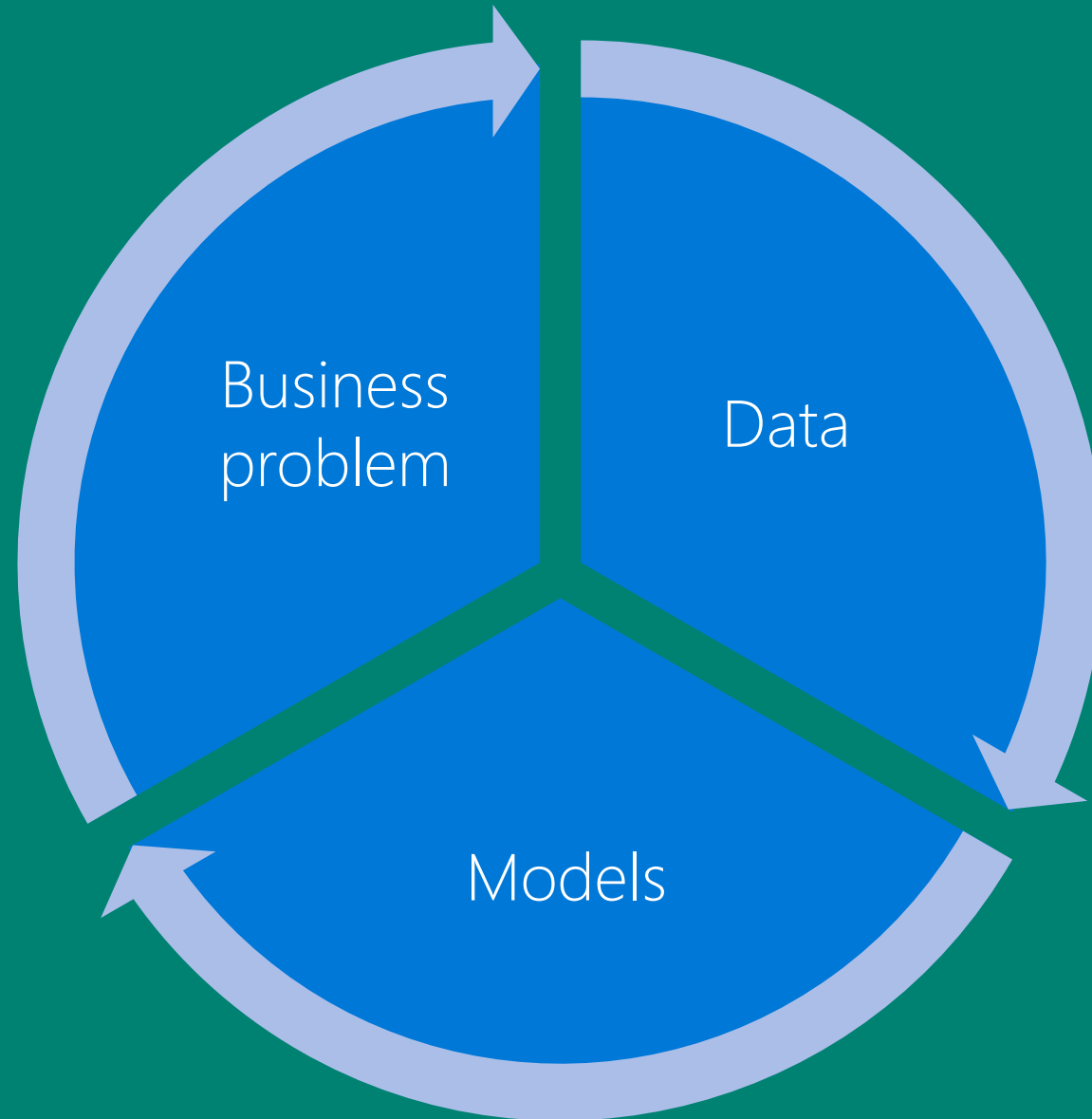


Data science

Scientific method of reasoning applied to data-driven decisions

Hypothesis,
experiments, facts,
logical reasoning
+ data engineering.

How?



Machine learning \equiv data mining

Actually: data mining uses machine learning.

Explores
data

Finds
patterns

Predicts
(scoring)



The Periodic Table of Data Science

An overview of key companies, resources and tools in data science (as of 4/12/2017)



Dc DataCamp	Ga General Assembly	Sd Strata Data
Sb SpringBoard	M Metis	Od ODSC
Ex Edx	Di Data Incubator	Tc Tableau Conference
C Coursera	In Insight	U UseR!
Uda Udacity	Dsa NYC Data Science Academy	Pd PyData
Ude Udemy	G Galvanize	Paw Predictive Analytics World
Ps Pluralsight	Dsg Data Science for Social Good	Kdd ACM SIGKDD Conference
Ly Lynda	Dsy Data society	Tpc Teradata Partners Conference
Tt TeamTreeHouse	Dsj Data Science Dojo	Icd IEEE International Conference on Data Mining
Bdu Big Data University		

Courses
Boot camps
Conferences

Data
Projects & Challenges, Competitions
Programming Languages & Distributions

Search & Data Management
Machine Learning & Stats
Data Visualization & Reporting

Collaboration
Community & Q&A

News, Newsletters & Blogs
Podcasts

Py Python	Js JavaScript	Vb Visual Basic	Pgs PostgreSQL	Sql SQLite	Ah Apache Hadoop	W Weka	Bml BigML	Kn Knome	Sm Spark MLlib	Pb Power BI	Obi Oracle BI	Shn Shiny	Ddl Domino Data Lab	De Data Science Experience
R R	Cp C++	Sc Scala	Ar Amazon Redshift	Bq Google BigQuery	Hw Hortonworks	O Oracle	Dar DataRobot	Lib LibSVM	Ho H2O	Bo BusinessObjects	Alt Alteryx	Mpl Matplotlib	Nt Nteract	Rs Rstudio
S SQL	Pl Perl	Ca Cassandra	Hb HBase	Td Teradata	Cl Cloudera	Mss Microsoft SQL server	Rm RapidMiner	Mat Mathematica	Th Theano	Sp Spotfire	Sav SAS Visual Analytics	Ply Plotly	Ro Rodeo	Be Beaker Notebook
B Bash	Mr Microsoft R Open	P Pig	Mdb Mongo DB	To Toad	Aem Amazon Elastic Mapreduce	Spl Splunk	Cho Chorus	Mah Mahout	Aml Azure Machine Learning	Ql Qlikview	Po PowerPivot	Me Microsoft Excel	Spy Spyder	Ze Apache Zeppelin
Mtl Matlab	Cy Canopy	Im Impala	K Kafka	Ms MySQL	Mar MapR	Sr Solr	Tf Tensorflow	St Stata	D D3	Co Cognos	Gch Google Charts	Pe Pentaho	Dst Data Science Studio	Ju Jupyter
J Java	An Anaconda	Sp Spark	Hi Hive	Idb IBM DB2	Lu Lucene	El ElasticSearch	Sk Scikit-Learn	Da Dato/Graphlab	My Microstrategy	Aa Adobe Analytics	T Tableau	B Bokeh	Db Databricks notebook	Gh Github

Dw Data.world	Q Quandl	Fte FiveThirtyEight	Sa Socrata	Gp Google Public	Dg Data.gov	K Kaggle
St Statista	Uci UCI Machine Learning Repository	Wb World Bank	At Academic Torrents	Bf Buzzfeed	Dk DataKind	Dd DrivenData


Re Reddit	So Stack Overflow	Cv Cross Validated	Qu Quora	Av Analytics Vidhya	Dse Data Science Stack Exchange
Mu Meetup	Rdm RDataMining				

Kdn KDnuggets	Ibd insideBIGDATA
Rb R-Bloggers	Pp PlanetPython
Hn HackerNews	Dt DataTau
Dsc Data Science Central	Dsr Data Science Roundup
Dsw Data Science Weekly	Or O'Reilly
Dr Data Elixir	Pw Python Weekly
Rw R Weekly	Pd Partially Derivative
Bds Becoming a Data Scientist	Tm Talking Machines
Ds Data Stories	Dsk Data Skeptic
Ld Linear Digressions	Ns Not So Standard Deviations



Chart from "The Periodic Table of Data Science" 2017, DataCamp

Suggested approach



Visualizing
Data

Easy, visual,
intuitive, Excel,
just works

R

Descriptive stats,
feel your data,
more algorithms

Azure ML

Cloud service,
more algorithms,
auto-tuning,
no more *throwing*
over the wall

Other Microsoft data science tools

HDInsight

Spark + Hadoop in the cloud
R Server
+ Storm, Kafka, HBase, Hive
+ etc.

Azure Stream Analytics

Streaming data originating in the cloud
Based on HDInsight/Hadoop

Azure Analysis Service

Microsoft SQL Server Analysis Services
in the cloud

Also useful:

Power BI: Power Query, Power View, and
Dashboards

Excel

Azure Data Factory (ETL in the cloud)

Analytics Platform System (SQL Server on
steroids + Hadoop + hardware)

Machine learning process and algorithms

How?

1. Define & initialise a model
2. Train model (process cases)
3. Validate model
...by scoring (making predictions) a test data set and evaluating the results
4. Use it: Explore or Deploy
...visualise and study
...deploy as a (web) service
5. Update and revalidate

Algorithm classes

Classifiers	Predict what class case belongs to
Clustering	Discover natural groupings of cases
Regression	Predict numerical outcomes
Recommenders	Explore associations between cases
Ensembles	mix them up

Machine learning and stats

with R, Notebooks

What is R?

Language, interpreter, *poor* IDE

5000+ packages of statistical software

Basic IDE: RStudio

<http://www.rstudio.com/>

Visual Studio (makes it even easier)

Open source, free, multiplatform

Core R: the purest version: <http://cran.r-project.org/>

Microsoft R Open, formerly known as Revolution R Open (RRO): <https://mran.microsoft.com/open/>

Azure ML: built-in

Notebooks

Develop experiments

Jupyter and Zeppelin (most popular notebooks)

Azure ML: built-in

Azure Notebooks (Jupyter)

Languages support:

Python 2, Python 3, R, F#

Machine learning

with Azure ML

Azure ML

Machine Learning platform in Azure cloud

Pre-process data

Engineer features

Modelling \equiv machine learning \equiv data mining

Run R

Run Python

Experiments (modelling) + Web Services (deployment)

Free

Limited: data size, experiment duration, scalability, speed

Paid

Relatively inexpensive, can be free

Demo

Tornadoes-loss

(<http://tornadoes.azurewebsites.net/>)

Data



Azure Machine Learning



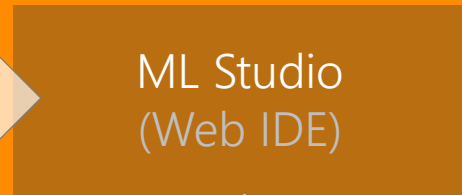
&

Consumers

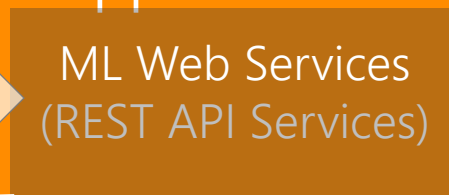
Azure Web Apps



Data



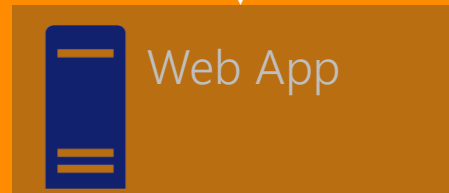
Model



API



Manage



API



Data

Modeling

Web App

Clients

Demo
create model + experiment

Demo

create web service

Demo develop web clients

Please evaluate this session
Your feedback is important to us!

Visit

<https://aka.ms/cc9cf1>



Q + A

&

Thank U