HEADS

PhD PROGRAMME IN HEALTH DATA SCIENCE

Machine Learning and Data Mining

Start your engines!

Basics of Health Intelligent Data Analysis

PhD Programme in Health Data Science

Pedro Pereira Rodrigues

U.PORTO
FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

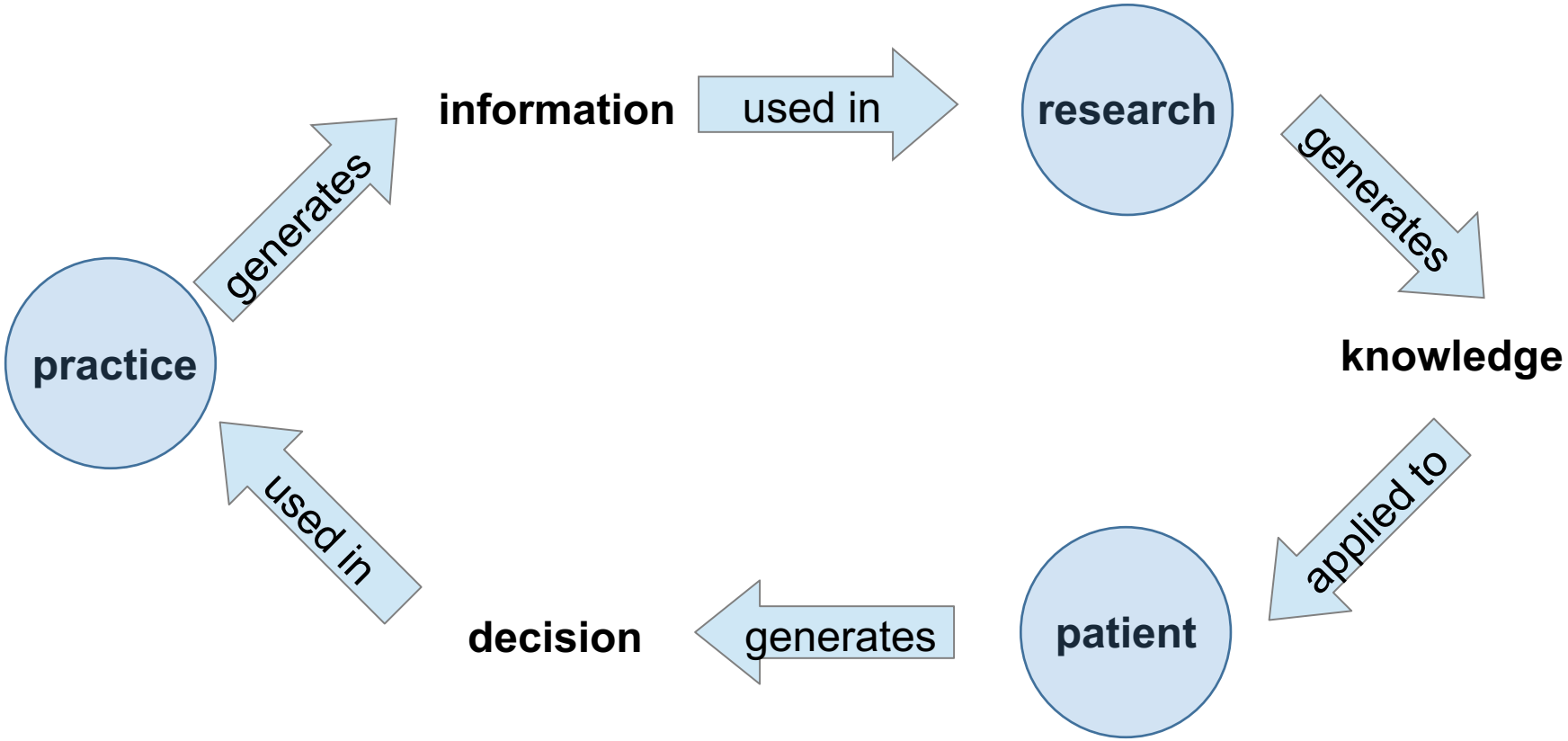
MEDCIDS
DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE

CINTESIS
Health. Research.

Evidence Based Medicine

"Conscient, explicit and criterious use of the best available evidence in clinical decision"

Sackett D. (1996)


practice
generates
information
used in
research
generates
knowledge
applied to
patient
generates
decision
used in

Real-World Biomedical Data

"The complicated nature of real-world biomedical data has made it necessary to look beyond traditional biostatistics."

Lucas P. (2004)

Wealth of Health Data

"The routine operation of modern healthcare systems produces a wealth of data in electronic health records, administrative databases, clinical registries, and other clinical systems."

Peek & Rodrigues (2018)

The 42 V's of Big Data and Data Science

Tom Shafer

April 1, 2017

Tweet Share Like 49 Share G+

https://www.elderresearch.com/company/blog/42-v-of-big-data

Knowledge Discovery

"It is widely acknowledged that there is great potential for utilizing these routine data for health research to derive new knowledge about health, disease, and treatments."

Peek & Rodrigues (2018)

Computational Statistics

"Computational statistics is a branch of mathematical sciences concerned with efficient methods for obtaining numerical solutions to statistically formulated problems."

Nickel C. (2020)

Data Science

"Study on creation, validation and transformation of data to generate meaning."

Data Science Association (2020)

Data Science Process

"Data science process begins with asking an interesting research question that guides the overall workflow of the data science project."

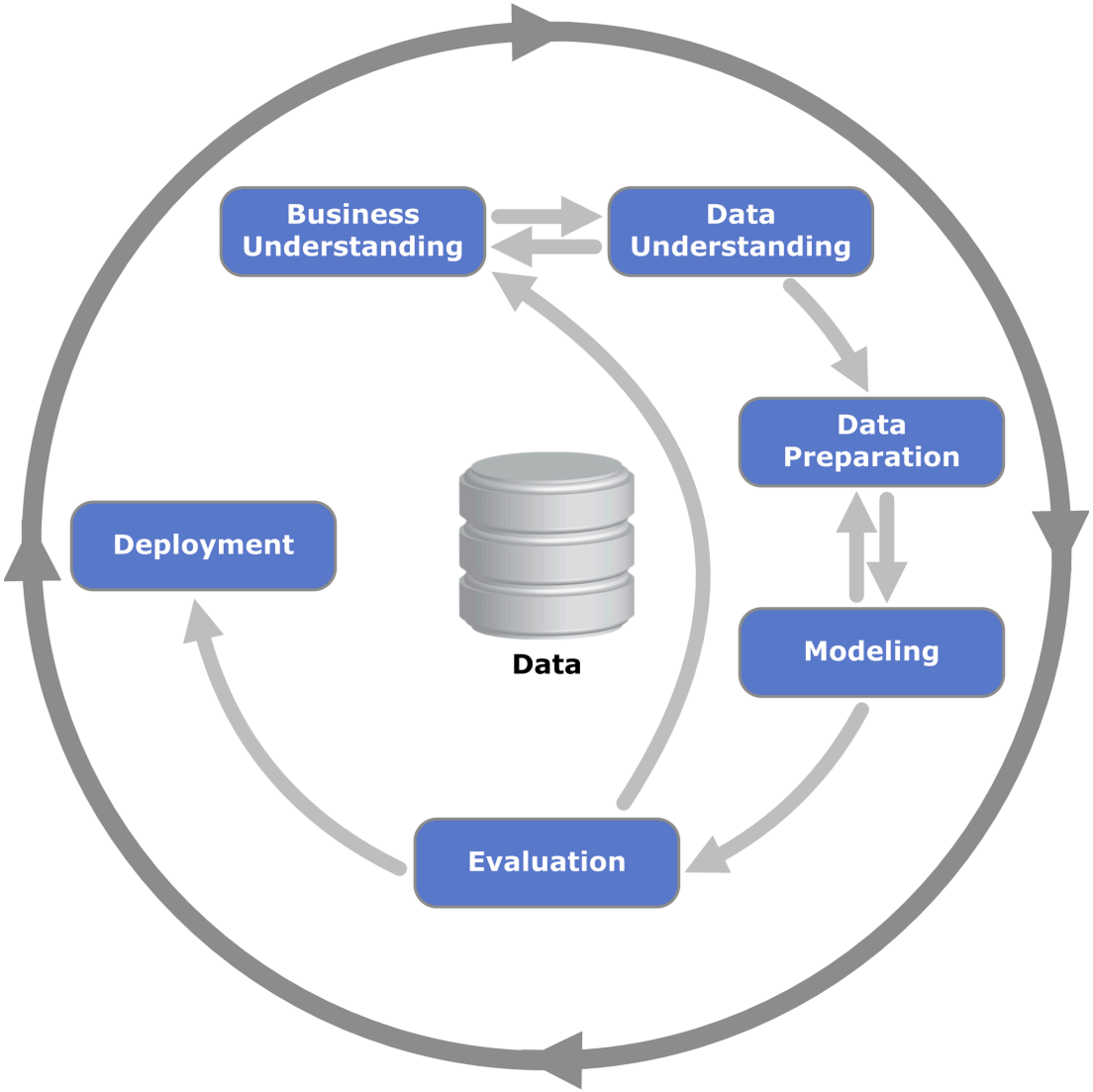
DeZyre (2016)

Data Science Workflow

"Having a well-defined workflow for any data science project is less frustrating for any data professional to work on."

DeZyre (2016)

Data Mining


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

Data Science Lifecycle

"The lifecycle of a data science project is not definitive and can be altered accordingly to improve the efficiency of a specific data science project as per the research requirements."

DeZyre (2016)

Clinical Knowledge Representation

"Clinical cases are getting more and more complex, yielding the application of modelling techniques likewise increasingly complex."

Lucas P. (2014)

Computational Intelligence

"For a computer to be intelligent, it has to be programmed appropriately. Ideally you would like to tell it only as much as it needs to know in a high-level language"

Poole D. (1998)

Artificial Intelligence

"Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension."

EU Commission (2019)

Artificial Intelligence

"AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions."

EU Commission (2019)

Artificial Intelligence

"As a scientific discipline, AI includes several approaches and techniques, such as machine learning, machine reasoning, and robotics."

EU Commission (2019)

Machine Learning

"The field of machine learning is concerned with question of how to construct computer programs that automatically improve with experience"

Mitchell (1997)

Supervised Machine Learning Metaphor

"There is a teacher who teaches the system a concept, with which the student is able to classify new cases, and there is an error function for that classification."

Hastie T., Tibshirani R. & Friedman J. (2001)

Inductive Bias

"A learner that makes no a priori assumptions regarding the identity of the target concept has no rational basis for classifying any unseen instances."

Mitchell (1997)

Model Performance

"The generalization performance of a learning method relates to its prediction capability on independent test data."

Hastie T., Tibshirani R. & Friedman J. (2001)

Black Boxes

"Some machine learning techniques, although very successful from the accuracy point of view, are very opaque in terms of understanding how they make decisions."

EU Commission (2019)

A course on machine learning and data mining

Aims to empower students with the necessary knowledge and skills to:

- interpret and apply machine learning techniques in health databases
- identify problems that can be addressed with data mining processes
- recognize the most common tasks of knowledge discovery (e.g. clustering, classification, association, regression)
- apply and interpret the obtained results, according to technical accuracy and impact in the domain.

An introduction to the statistical programming language R will be presented as part of the course and students will be required to complete their assignments in R.

A course on machine learning and data mining

INTRODUCTION

CLASSIFICATION & VALIDATION

includes cross-validation

DECISION TREES

includes random forests

BAYESIAN NETWORKS

includes classifiers

NEURAL NETWORKS

includes deep learning

KERNEL METHODS

includes svms

ENSEMBLE MODELS

includes bagging and boosting

BIG DATA - BIG MODELS

includes streams

ASSOCIATION RULES

includes apriori

CLUSTER ANALYSIS

includes hierarchical clustering

ANOMALY DETECTION

includes outliers and rare events

DATA PREPROCESSING

includes feature engineering

TEXT MINING

includes nlp

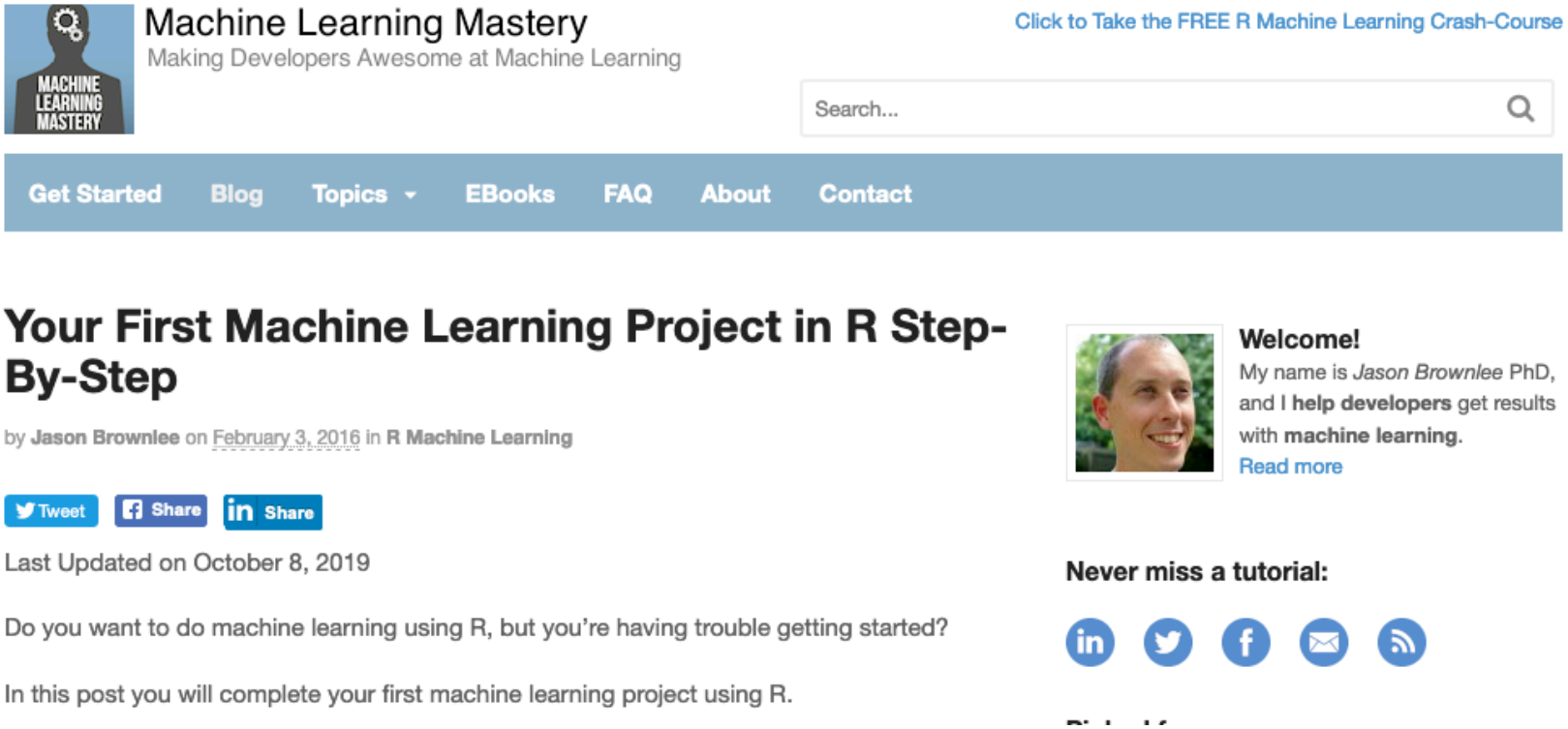
VISUAL DATA MINING

includes geospatial analysis

Software Tools for Machine Learning

Software Tools for Machine Learning

Your First Machine Learning Project in R



https://machinelearningmastery.com/machine-learning-in-r-step-by-step/