Hes·so
Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences
Western Switzerland

MSE | MASTER OF SCIENCE IN ENGINEERING

Multimodal Processing, Recognition and Interaction

Predictive Analytics for Temporal Forecasting

Simon Ruffieux

Elena Mugellini, Stefano Carrino, Omar Abou Khaled

Introduction

- Predictive Analytics

Broad term describing a variety of **statistical and analytical techniques** used to **develop models** that **predict future** events, state or behaviors

- Data mining

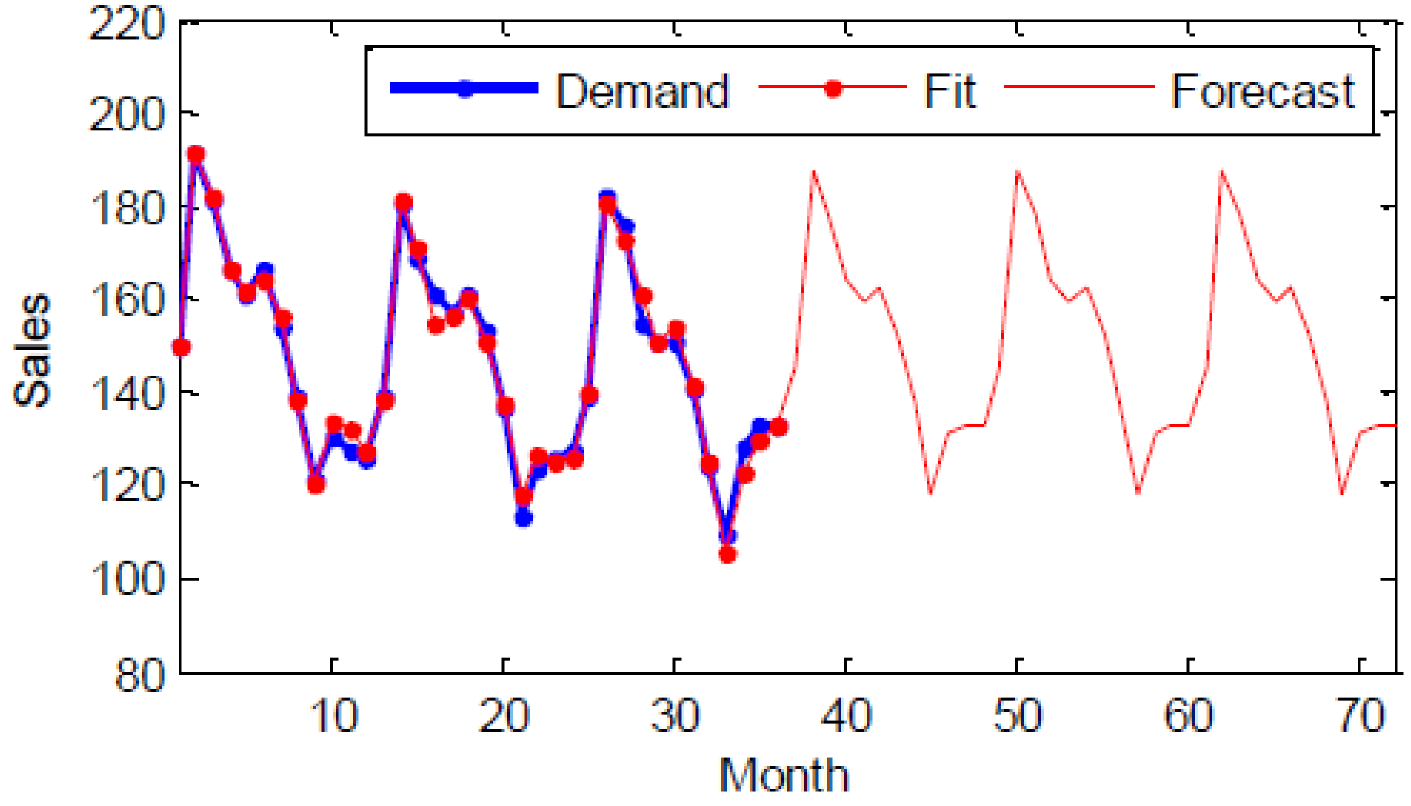
Component of predictive analytics that entails analysis of data to identify **trends**, **patterns**, or **relationships** among the data

→ Data mining is used to create the predictive model

Introduction

- Temporal forecasting

A predictive model that is able to use current and past observations to **predict future** observations



Introduction

- Important characteristics to consider for forecasting **time-series**
  - Is there a **trend**, meaning that, on average, the measurements tend to increase (or decrease) over time?
  - Is there **seasonality**, meaning that there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?
  - Are their **outliers**? In regression, outliers are far away from your line. With time series data, your outliers are far away from your other data.
  - Is there a **long-run cycle** or period unrelated to seasonality factors?
  - Is there **constant variance** over time, or is the variance non-constant?
  - Are there any **abrupt changes** to either the level of the series or the variance?

Introduction

- Forecasting time-series
  - **Univariate** (one variable) forecasting:
    - ✓ Forecasts Y from trend alone
  - **Multivariate** (many variable) forecasting:
    - ✓ Forecasts Y from trend and other variables $X_1, X_2, \dots$
    - ✓ Allows for “what if” scenario forecasting
    - ✓ May or may not make more accurate forecasts

Introduction

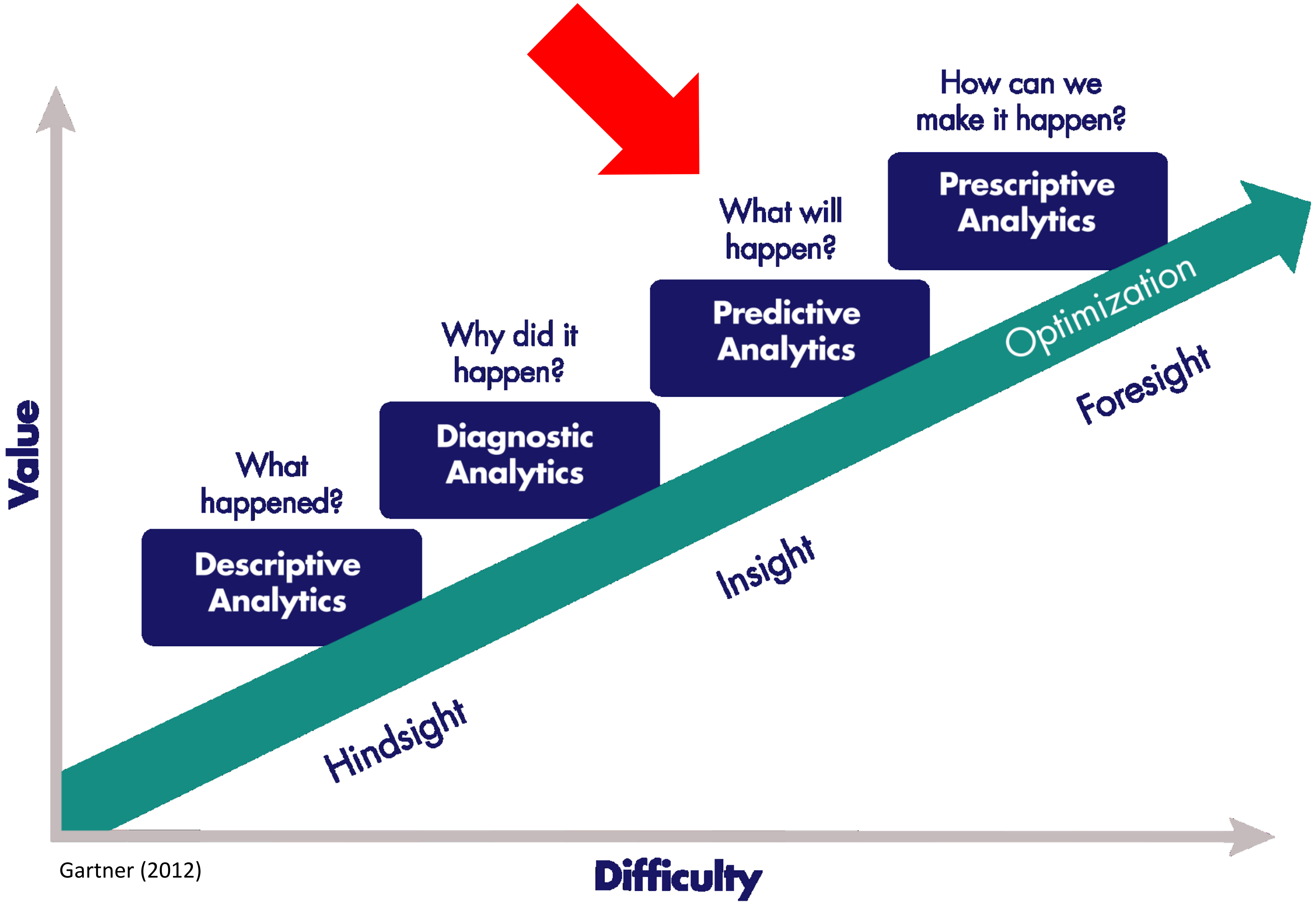
- What is the difference with classification or regression methods (see next lectures) ?
  - Almost NONE -> both can be applied to find the predictive model. The chosen method depends on the problem

Use algorithms to find a model

- **Predictive model**

Observations until time t (the input) are used to **forecast future values** at time t + k, with k > 0 (the output)

Types of Analytics



Gartner (2012)

Problem outline - Types of Analytics

- **Descriptive analytics**

  Describe the data with counts and averages and report as table, pie charts, etc.

- **Diagnostic analytics**

  Examine the data and try to find explanations using drill-drown techniques, data discovery and correlations

- **Predictive analytics**

  By learning what happened in the past, forecast what might happen in the future.

- **Prescriptive analytics**

  Examine data in order to provide recommendations to answer the question “What can be done to make x happen ?”

What Predictive Analytics Cannot Do

“The purpose of predictive analytics is NOT to tell you what will happen in the future. It cannot do that. In fact, no analytics can do that. Predictive analytics can only forecast what might happen in the future, because all predictive analytics are probabilistic in nature.”

(Bertolucci, 2013)

Example of domains of Application

- User intent
  - J. Kiseleva, H.T. Lam, M. Pechenizkiy, T. Calders, “Predicting current user intent with contextual markov models” ICDMW, Workshop Proc. of the IEEE 13th Int. Conf. on Data Mining, 2013, pp. 391-398.
- Food demand/Food sales prediction
  - I. Žliobaitė, J. Bakker, M. Pechenizkiy, (2012) Beating the baseline prediction in food sales: How intelligent an intelligent predictor is? Expert Systems with Applications 39 (1), pp. 806-815.
- FinTech – Stock prediction
  - See examples
- Healthcare
  - “It's predicted that, in the coming decade, developers will have a greater impact on the future of healthcare than doctors.”
  - Sick care (current) vs Health care (future) !
- Predictive Maintenance/ fault prediction

Main Challenges of Forecasting Timeseries !

- Problem are often dynamic (notion of <**concept drift**>)
  - Evolve over-time
- Need to understand data and its evolution (<**Trend and Seasonality**>)
  - Weekly patterns ?
  - Daily patterns ?
  - How much past data to take into accounts ?
  - External factors ?
    - Unpredictable factors …
- Practical use in real-conditions
  - Long-term predictions harder to score
  - Long-term predictions might be influenced by actions based on short-term predictions

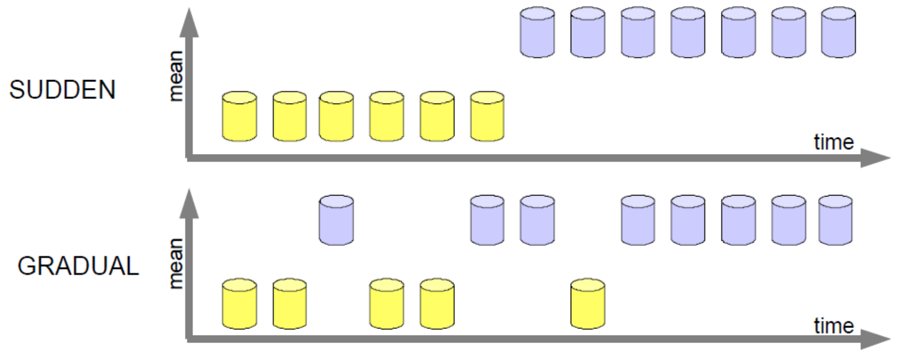
Concept drift (I) - Definition

In the real world, **concepts are often not stable but change over time**

This implies that a model built on old data becomes inconsistent with the new data, and **regular updating** of the model is necessary

→ Often the cause of change is hidden, not known a priori, making the learning task more complicated

- Two types of concept drifts in litterature
  - Sudden
  - Gradual



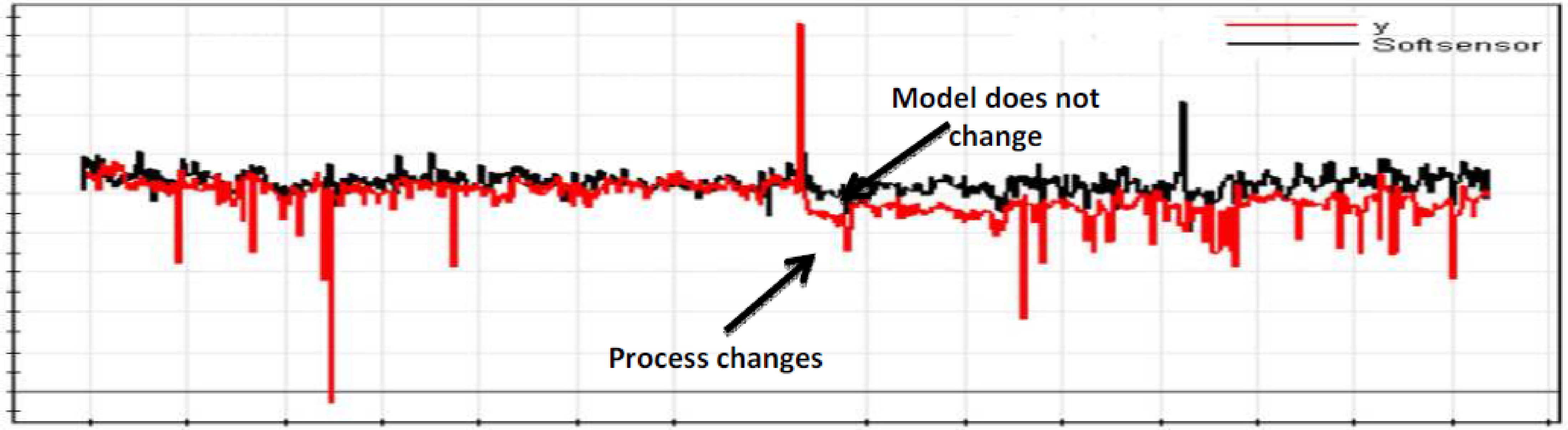
REF- http://www-ai.cs.uni-dortmund.de/LEHRE/FACHPROJEKT/SS12/paper/concept-drift/tsymbal2004.pdf

Concept drift (II) - Example

the supplier of raw material changes

a sensor breaks/ “wears off”/ is replaced

new operating crew



source: Evonik Industries

new regulations to save electricity

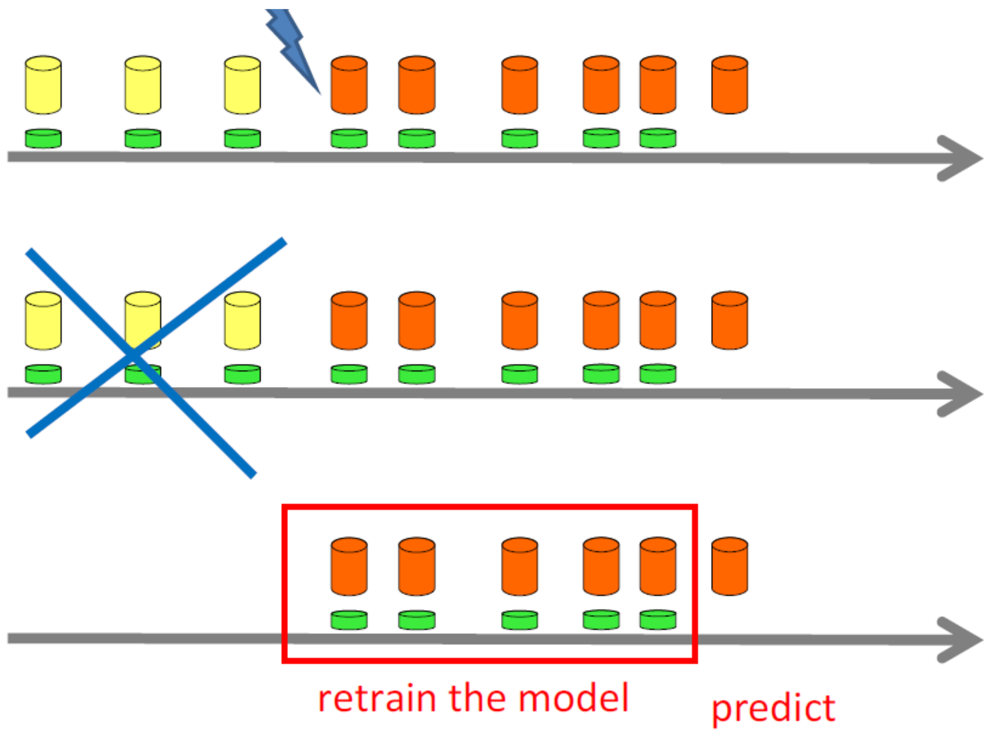
new production procedures

REF- https://www.cs.waikato.ac.nz/~abifet/PAKDD2011/PAKDD11Tutorial_Handling_Concept_Drift.pdf

Concept drift (III) – Handle drift

Adaptive learning strategies!

- Triggering
  - (1) detect change
  - (2) issue reaction



Detect change

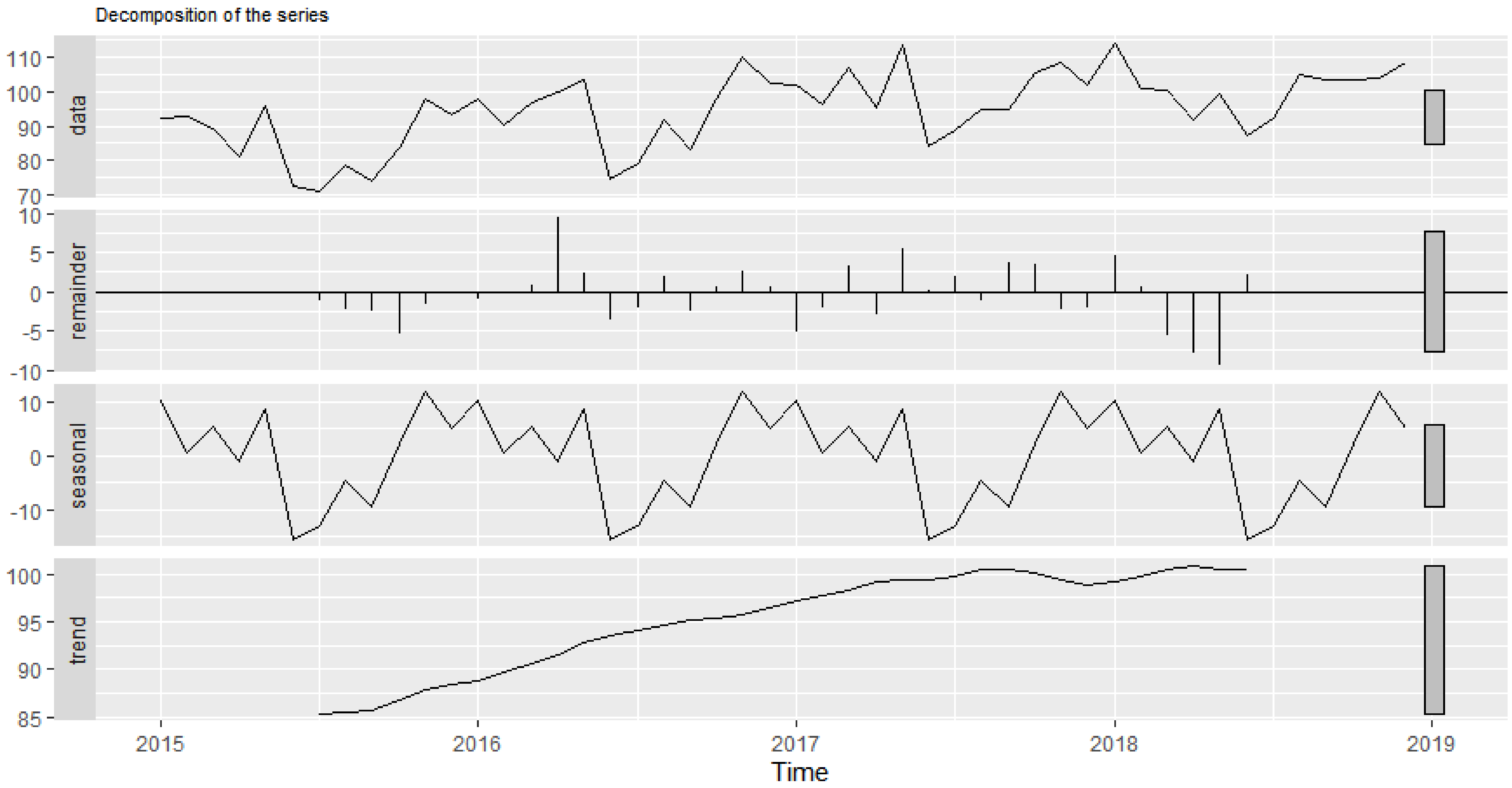
Ignore old data

Retrain on new data

- Evolving
  - Fixed windows (online learning)
  - Instance weighting
    - Example is LSTM algorithms (Long short term memory)

REF- https://www.cs.waikato.ac.nz/~abifet/PAKDD2011/PAKDD11Tutorial_Handling_Concept_Drift.pdf

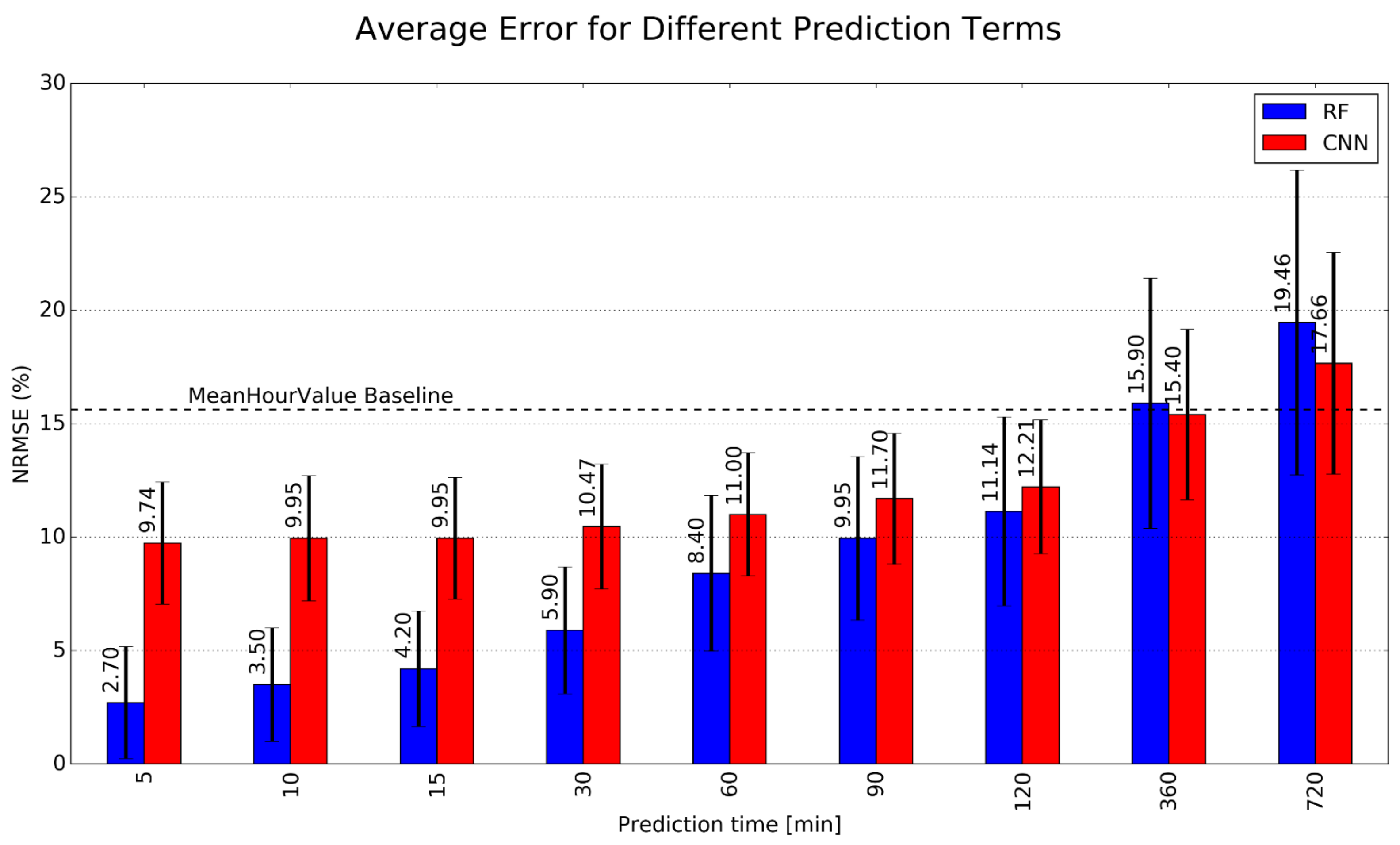
Concept of Trend and Seasonality



https://towardsdatascience.com/5-simples-steps-to-build-your-time-series-forecasting-model-62356336bc35

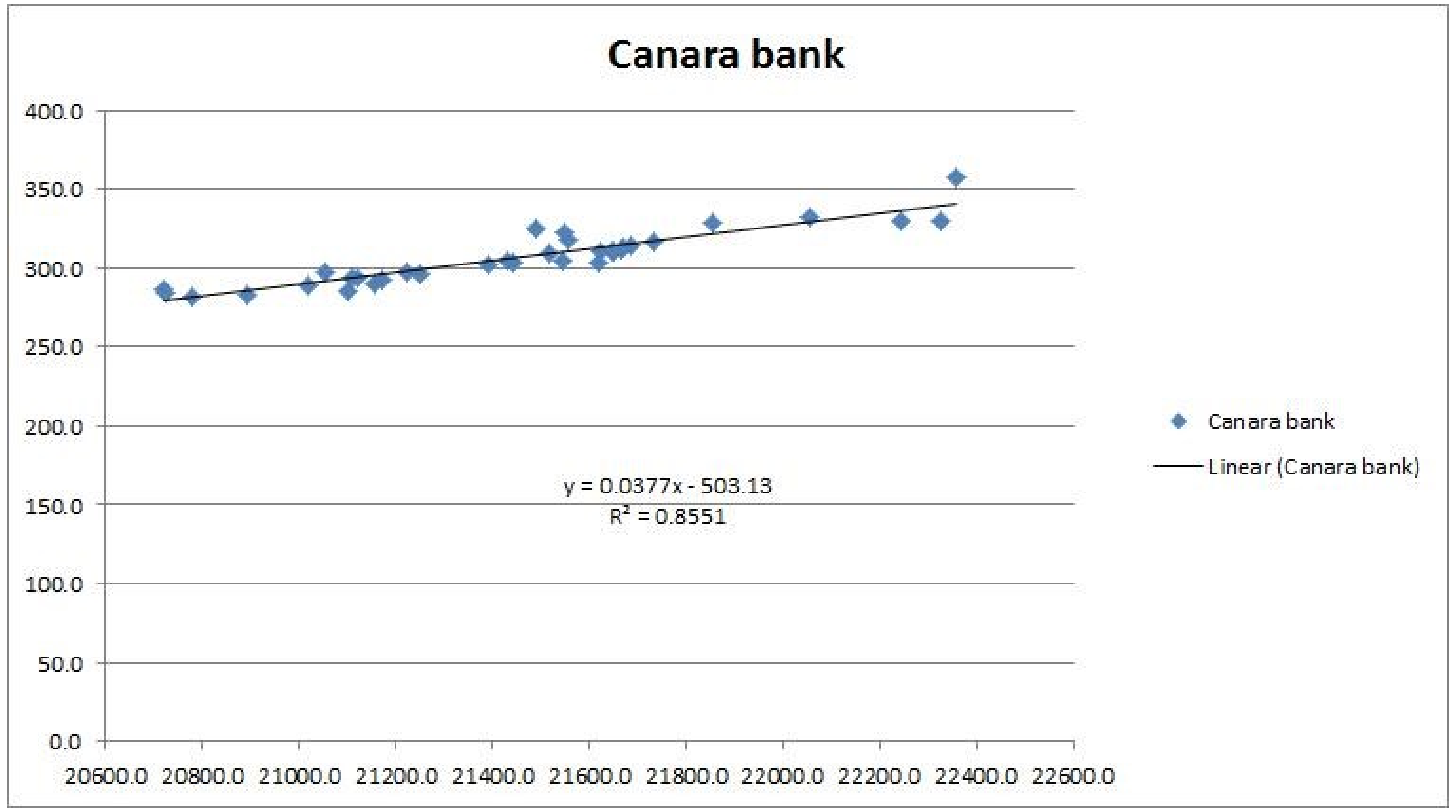
Challenge – Prediction horizon

- You might want to predict for different horizons!
  - Prediction accuracy usually decreases with increasing horizon
    - The «decrease» depends on system dynamics and changing rate



Examples – Stock predictions

- Predict stock using regressions
  - Very basic example
    - Try to predict Canara's bank stock using Bank Nifty's price

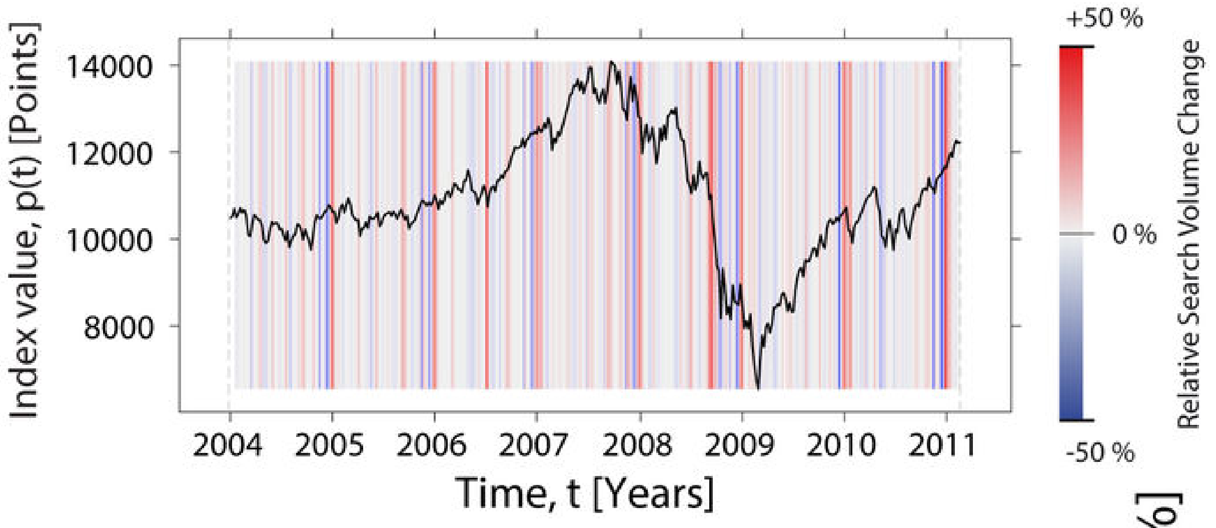


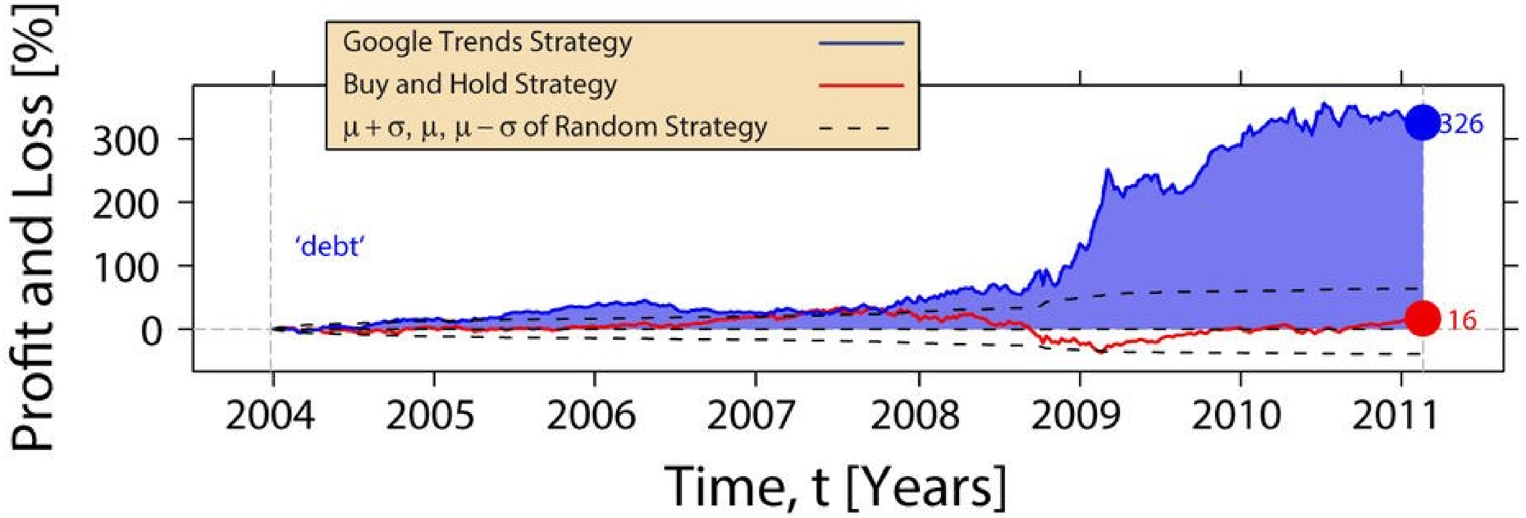
[REF] https://www.quantinsti.com/blog/machine-learning-trading-predict-stock-prices-regression/

Examples – Stock predictions

- Predict stock using Google Trends (Nature, 2013)

A recent study said it was now possible for computers to accurately predict whether stock prices will rise or fall based solely on whether there's an increase in Google searches for financial terms such as "debt." The idea is that investors get nervous before selling stocks and increase their Google searches of financial topics as a result. (https://www.nature.com/articles/srep01684)





Examples – Stock predictions

- Long-term Equity (stock) forecast (2016)
  - predict whether some company's value will be 10% higher or not over the period of one year

Table 3: *The results of machine learning based equity prediction using 10-fold cross validation and 11 selected features*

Algorithm	*Precision*	*Recall*	*F-score*
C4.5 decision trees	*0.660*	*0.660*	*0.660*
SVM with SMO	*0.636*	*0.629*	*0.624*
JRip	*0.640*	*0.639*	*0.639*
Random Tree	*0.700*	*0.700*	*0.700*
Random Forest	***0.765***	***0.765***	***0.765***
Logistic regression	*0.638*	*0.630*	*0.625*
Naïve Bayes	*0.526*	*0.515*	*0.453*
Bayesian Networks	*0.641*	*0.626*	*0.615*

They initially selected 28 features to perform predictions. Then they manually removed features until finding the optimal set of 11 features.

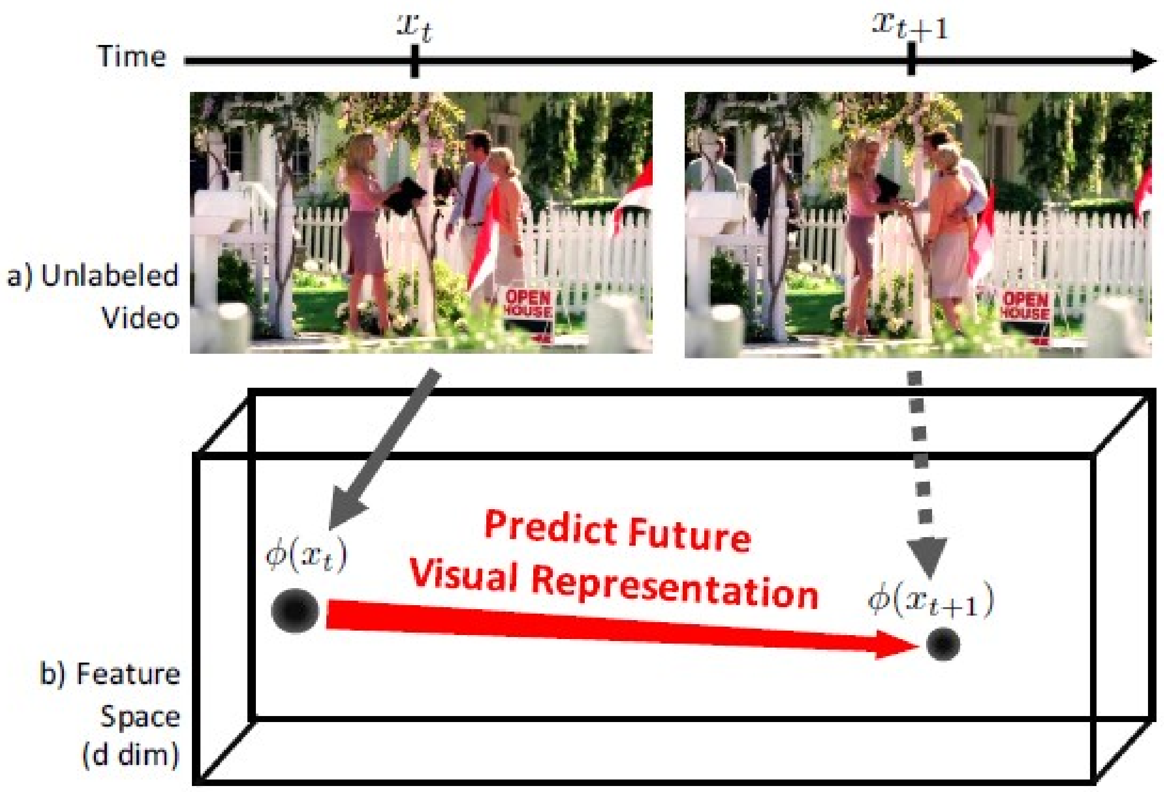
Equity forecast: Predicting long term stock price movement using machine learning, Journal of Economics Library, Nikola Milosevic (2016)

Examples – Predictive vision

- Predict next action in video stream
  - Deep-learning vision system anticipates human interactions using videos of TV shows
    - It uses a single frame in a video and tries to predict next action (5 second later)
      - ✓ hug, handshake, high-five, kiss (43% accuracy)

Video

https://youtu.be/AR3hY9iB5-I



- http://news.mit.edu/2016/teaching-machines-to-predict-the-future-0621

Examples – Votation prediction (I)

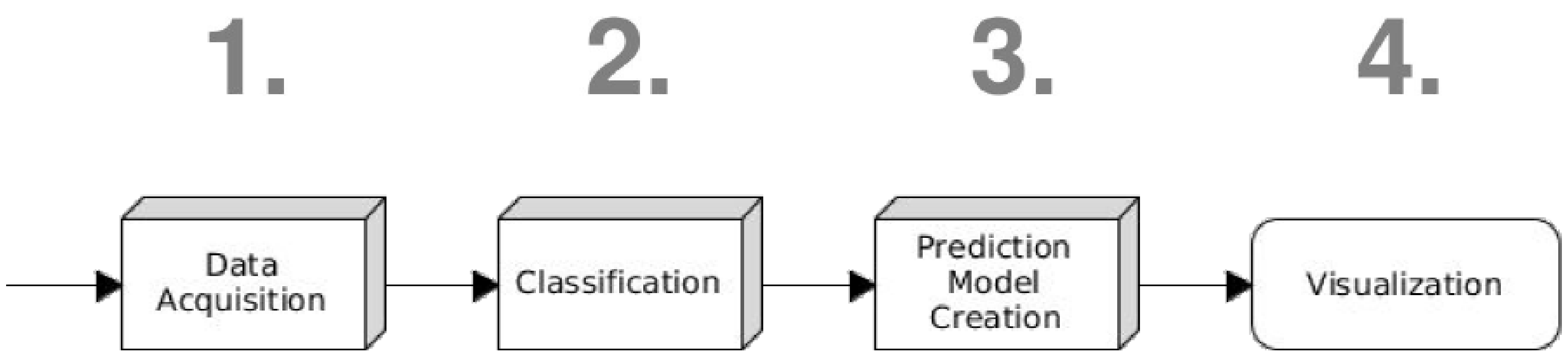
A Master Project by Jacky Casas

- Goal:
  - Predict the outcome of the votations of June 5th, 2016 (5 objects were submitted)
    - ✓ 1. Public Services
    - ✓ 2. Basic Income
    - ✓ 3. Transports
    - ✓ 4. Medically assisted reproduction
    - ✓ 5. Asylum

- How to do the predictions ?
  - Use twitter and newspaper data

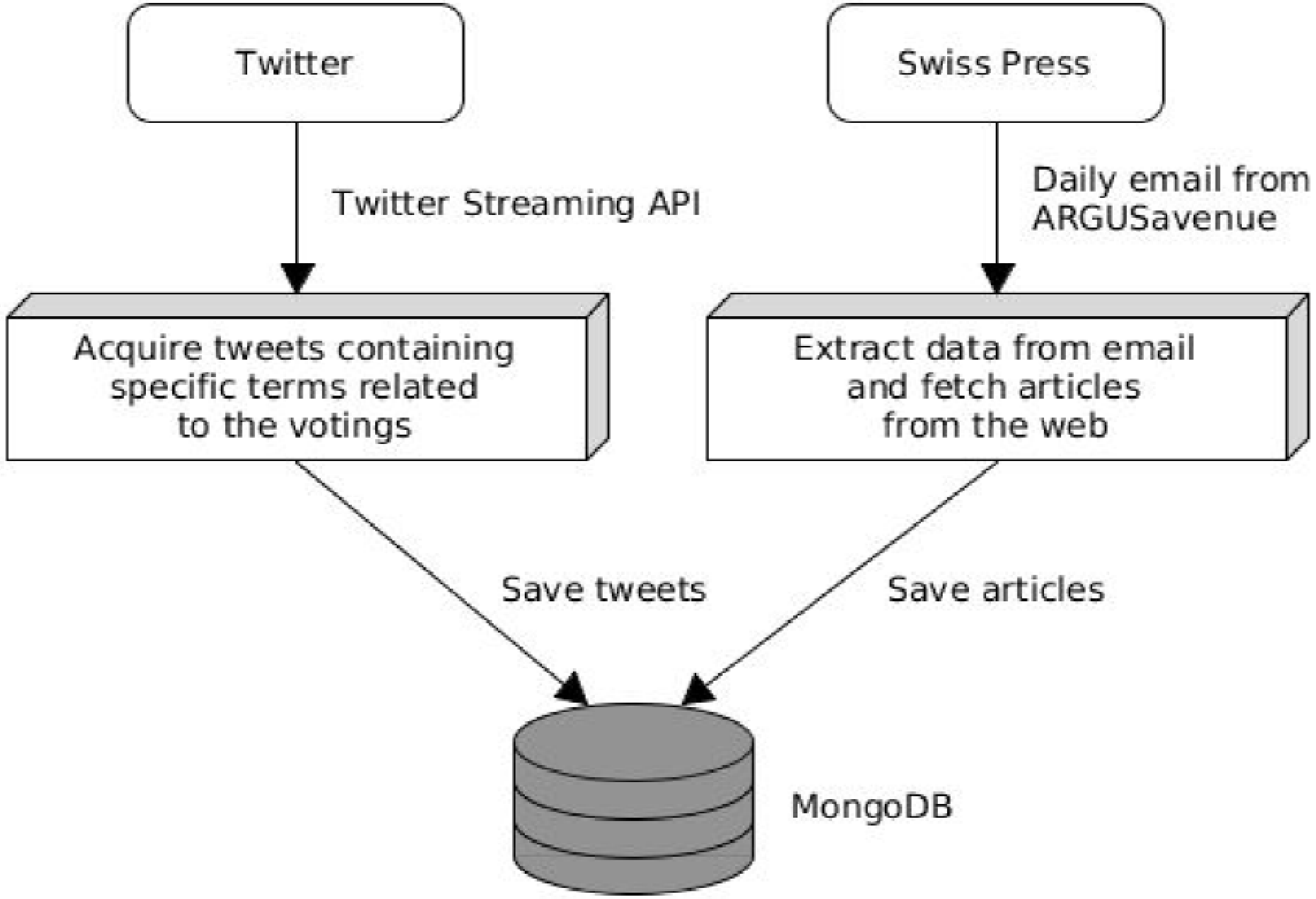
Examples – Votation prediction (II)

- The project involved the 4 following steps:



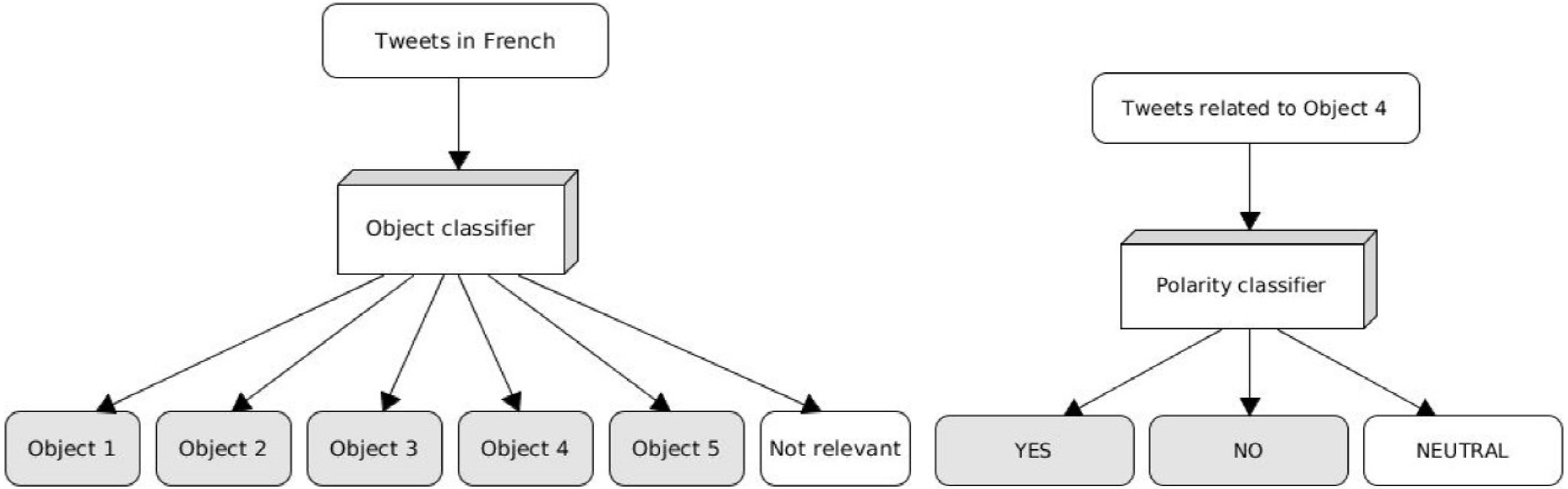
Examples – Votation prediction (III)

- **Data acquisition** from different sources



Examples – Votation prediction (III)

- Classification
  1. Classify tweets according to the votation object
  2. Classify polarity of a tweet for a specific object
     1. Positive, negative, neutral

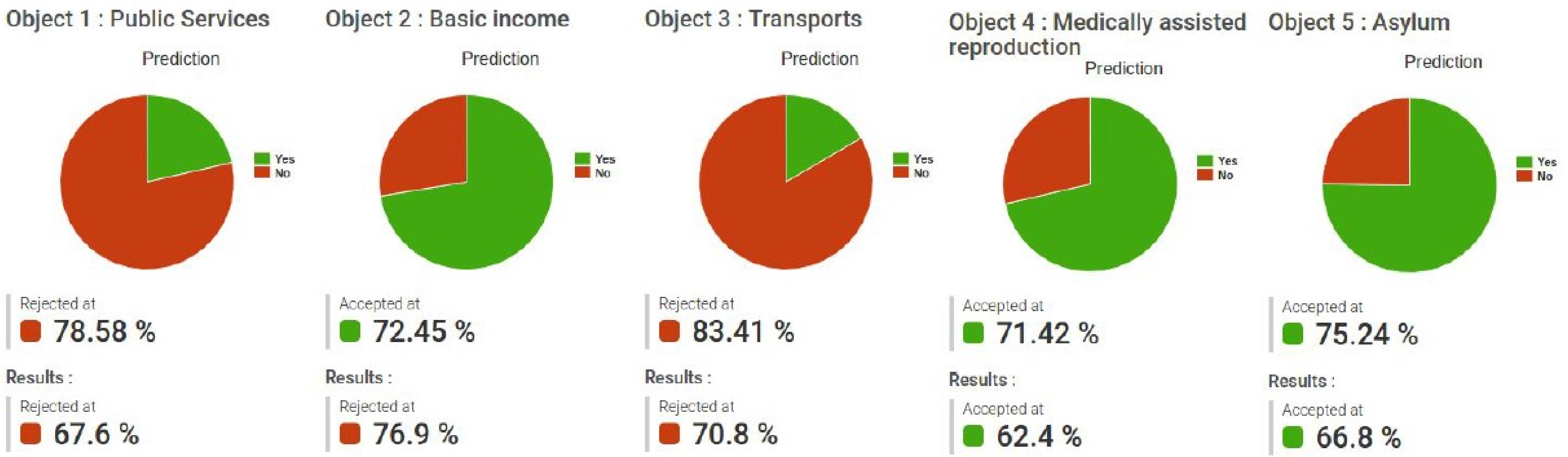


Examples – Votation prediction (IV)

- Prediction system
  - Based on potential visibility
    - The weights of a tweet is the quantity of followers of its author
    - Weights of all positive tweets are summed
    - Weights of all negative tweets are summed
  - The ratio positive/negative yields the prediction!

Examples – Votation prediction (V)

- Visualization & results



Mini Quiz

1. What is the type of input signal in Temporal Forecasting ?
2. What is a multivariate forecasting ?
3. Why do we have drift ?
4. What is the difference between seasonality and trend ?

Algorithm Recap – Random Forest

- What for ?
  - **Recognizing** the type (or **category**) of a scene captured in a photograph can be cast as classification, where the output is a discrete, categorical label (e.g. a beach scene, a cityscape, indoor etc.).
  - **Predicting** the **price** of a house as a function of its distance from a good school may be cast as a regression problem. In this case the desired output is a continuous variable.
  - **Detecting** abnormalities in a medical scan can be achieved by evaluating the scan under a learned probability density function for scans of healthy individuals.
  - Capturing the intrinsic variability of size and shape of patients brains in magnetic resonance images may be cast as **manifold learning**.
  - **Interactive image segmentation** may be cast as a semi- supervised problem, where the user's brush strokes define labelled data and the rest of image pixels provide already available unlabelled data.
  - **Learning a general rule** for **detecting tumors** in images using minimal amount of manual annotations is an active learning task, where expensive expert annotations can be optimally acquired in the most economical fashion.

Also known as CART (**Classification** and **Regression** Tree)

Following slides on Random Forest mostly inspired by "MSR Tutorial on decision forests by Criminisi et al. (2011)"

Algorithms – Random Forest

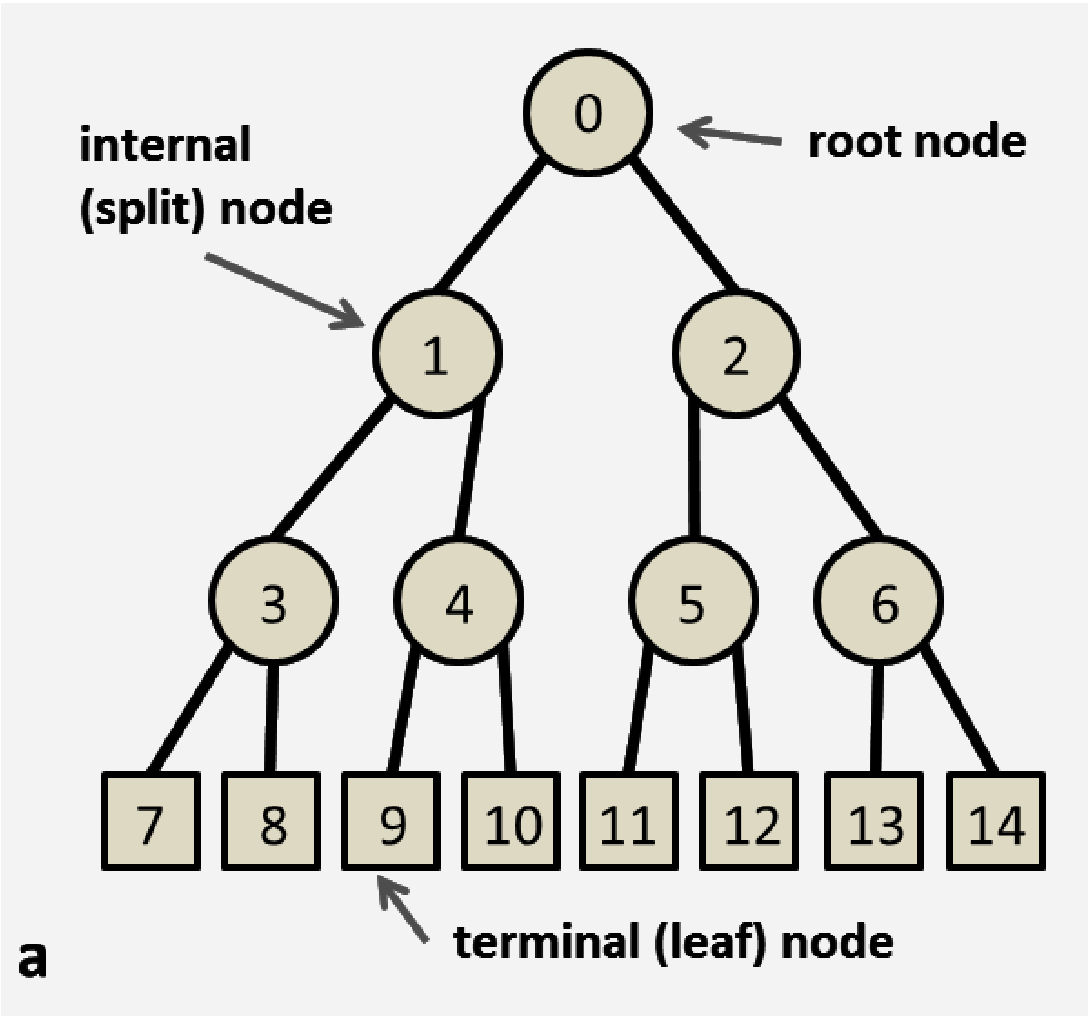
Recap: Decision Tree

A **tree** is a set of nodes and edges organized in a hierarchical fashion

- Root node
- Internal node (aka. split node)
- Edge
- Terminal node (aka leaf)

Each node has only 1 incoming edge but can have multiple outgoing edges

- Binary tree: 2 outgoing edges
- Ternary tree: 3 outgoing edges
- Heterogeneous tree : 2,3, x outgoing edges



An example of **general binary tree** structure

Algorithms – Random Forest

Recap: Decision Tree

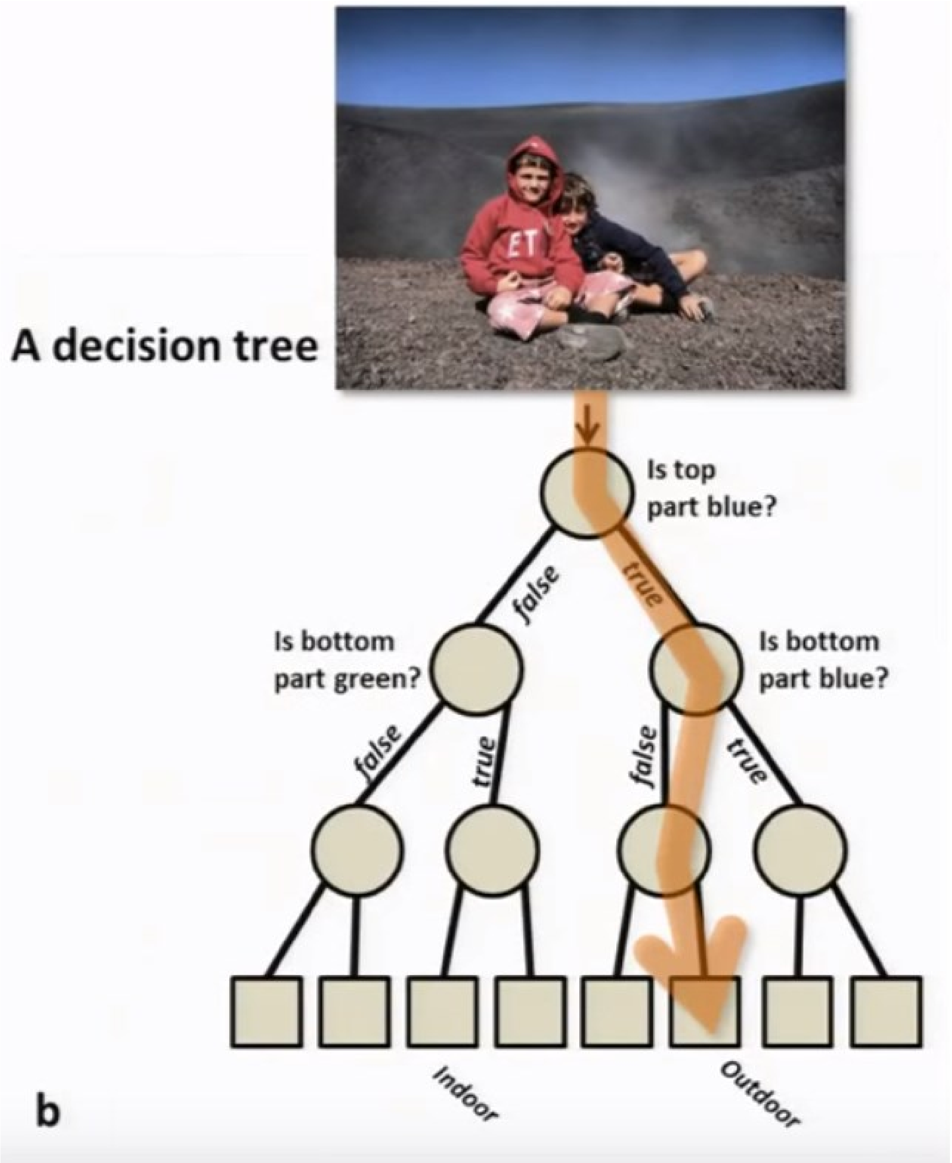
A decision tree is a binary tree where

- Each **split node** stores a test function for the incoming data
- A leaf may represents a final answer (**predictor**)

Example:

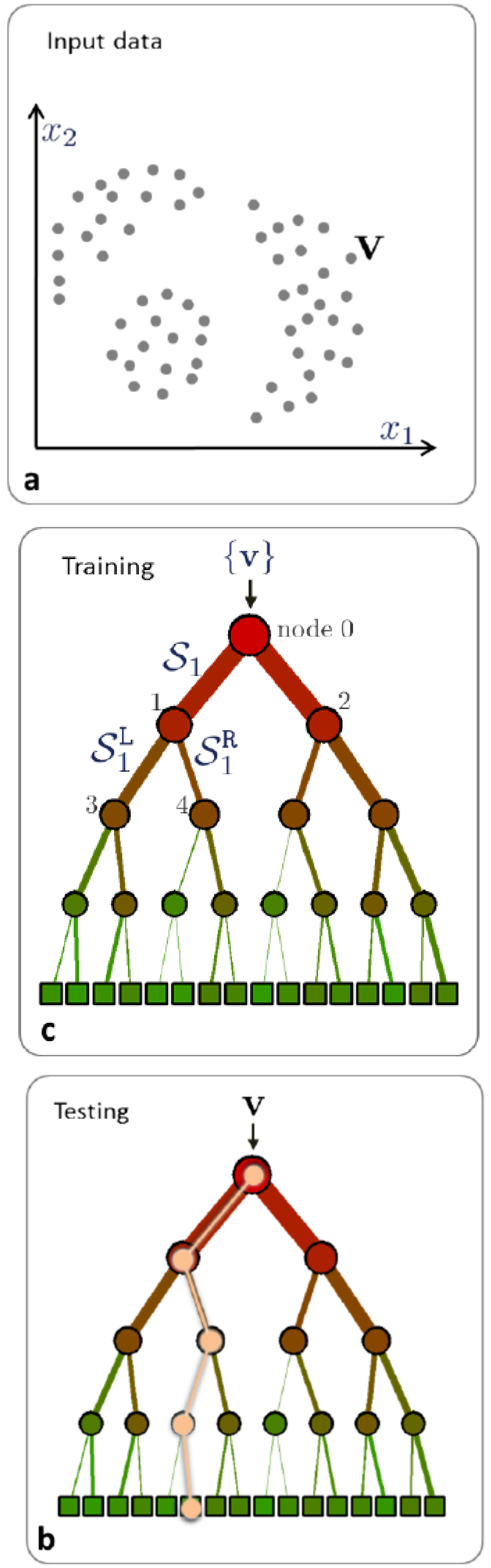
The decision tree shown on the right detects if a given picture represents an indoor or an outdoor scene.

A decision tree can be interpreted as a technique for splitting complex problems into a hierarchy of simpler ones.



An example of **decision tree** structure

Algorithms – Random Forest

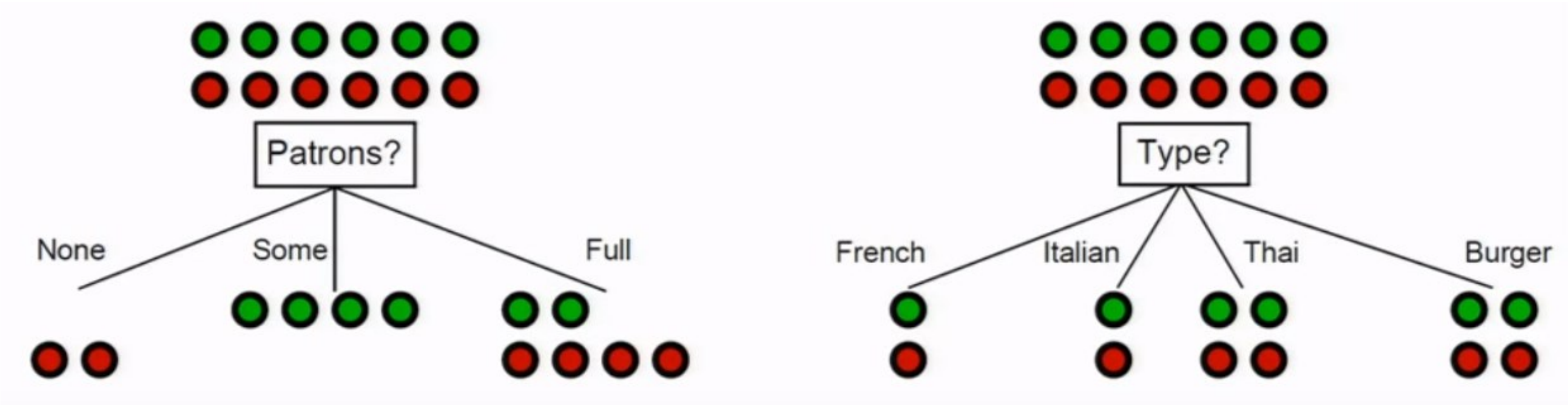


Recap: Decision Tree

To build the structure of the decision tree

- All training examples {v} are used
  - Optimizing the parameters of the split nodes so as to **optimize a chosen energy function**.

- Standard Energy function
  - Minimize Expected Entropy
  - Maximize Information gain (reduction in entropy) after each node

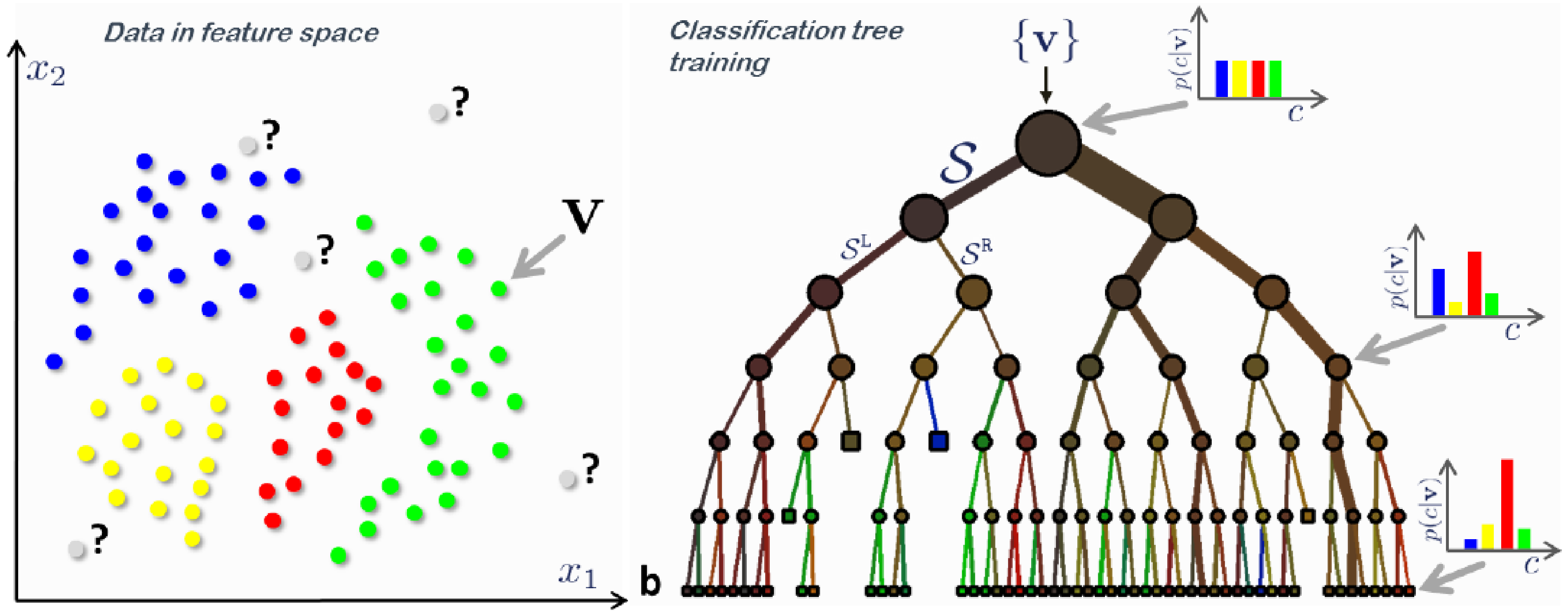


Which one is better ?

Algorithms – Random Forest

Recap: Decision Tree

It is not always possible to have pure terminal leaves !



Probabilities:
In the example above, the probability that a point v is RED is about 75% (5% blue, 10% yellow and 10% green). We assume the point v has always taken the right edge during Test procedure.

Algorithms – Random Forest

Random Forest

Instead of a single decision tree based on all the training data, build multiple decision trees each using a different subset of the training data and then fuse the results of each decision tree.

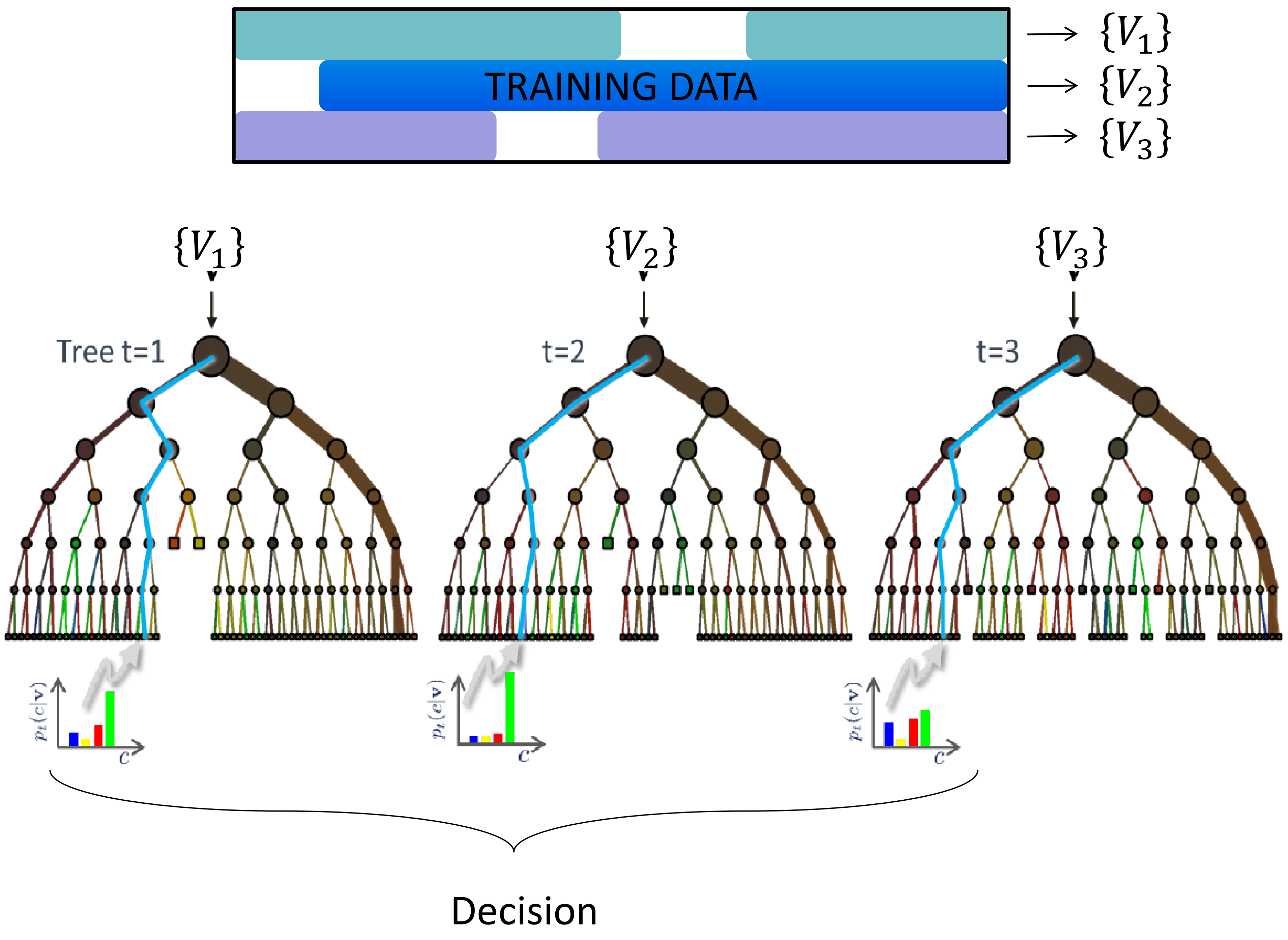
Training

1. Split training data randomly in n different subsets
   - → Obtain N training subsets
2. Use each subset to train a single Decision Tree
   - → Obtain N trained decision trees

Testing a sample

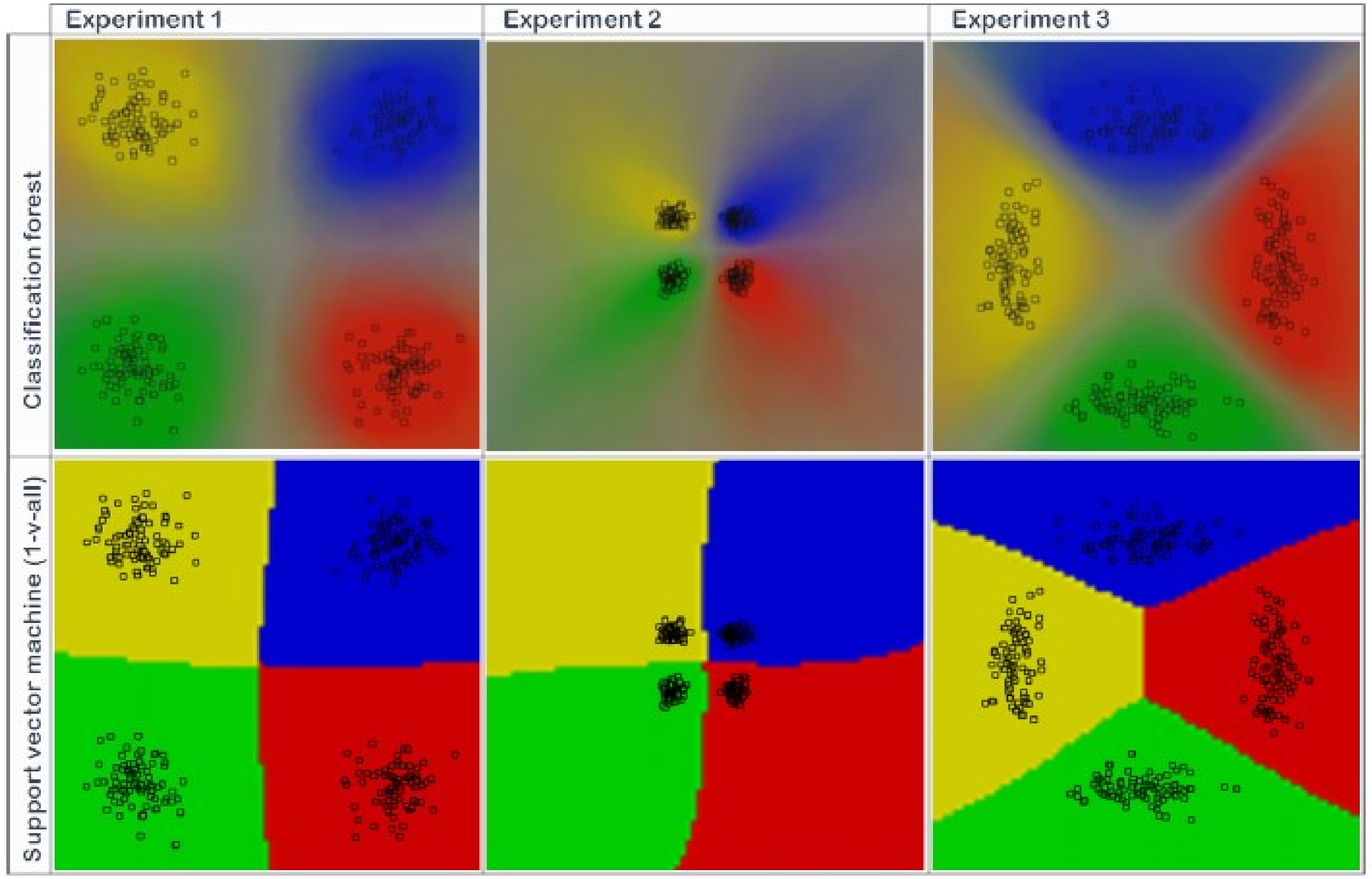
1. Test the sample using each decision tree
   - → Obtain N possible answer
2. Fuse the N answers into the most probable one

Algorithms – Random Forest



Algorithms – Random Forest

Random Forest vs SVM classification examples



A Random Forest provides **Confidence** information

Algorithms – Random Forest

- Advantages:
  - Provides **confidence** information
  - Provides **information** about each **feature importance (Gini coefficients)**
    - For minimal cost during training phase
  - Limited number of parameters
  - **“Fast” training**
  - Fully parallelizable
  - Resistant to outliers

- Disadvantages
  - Memory size of models
  - **“slow” prediction** (compared to other algorithms)
    - Highly dependent on # number of trees
  - Difficult interpretability of results

Algorithms – Random Forest – Scikit Parameters

- sklearn.ensemble.RandomForestClassifier
  - **n_estimators (100) :** the number of trees in the forest
  - **max_features ('auto')** : the size of the random subset of features to consider when splitting a node. (auto=sqrt(n_features))
  - **max_depth (None) :** the maximum depth of a tree
  - **min_samples_split (2) :** the minimum number of data points placed in a node before the node is split
  - **min_samples_leaf (1) :** the minimum number of data points allowed in a leaf node
  - **Bootstrap (true) :** Whether bootstrap samples are used when building trees. If False, the whole dataset is used to build each tree (see also *max_samples*)
  - **n_jobs (None) :** the number of jobs to run in parallel (-1 => use all processors)

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

https://scikit-learn.org/stable/modules/ensemble.html#parameters

Conclusion

- Temporal prediction is a subset of general machine learning problems
  - Warning - Ambiguity of the term prediction!
- Most ML algorithms can be used to perform temporal predictions
  - Good candidates: Random Forest, MLP, LSTM, SVM, etc.
  - Some algorithms have more adapted learning strategies (LSTM)
- Different prediction horizons for different purposes!
- Data evolves over time (Concept drift, Trend)
  - Sudden or Gradual
  - Different strategies exist to handle the drift

What You Should Know

- Predictive analytics
- Temporal predictions VS classical predictions
- Predictive vs other analytics
- Main challenges
  - Concept drift
    - Describe the problem
    - Provide solutions
  - Trend and Seasonality
- Examples of applications
- Explain Random Forest algorithm