PhD Program in «Industry 4.0»

An Explainable framework for Tourism Intelligence: the Apulian case

PhD candidate: Francesco De Nicolò

Tutor: Prof. Nicola Amoroso

Doctoral Research Seminars @ DEI, Politecnico di Bari – 9/2/2022

Politecnico di Bari

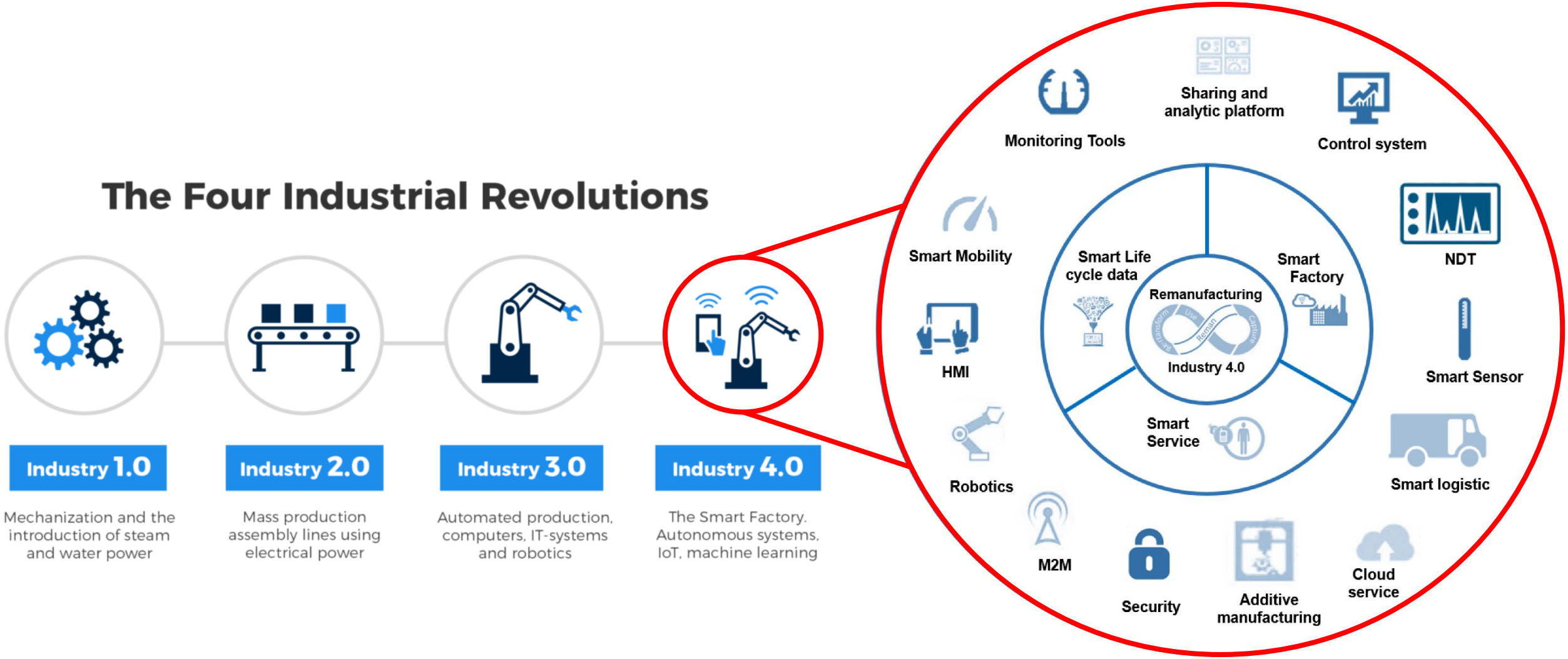
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Contents

- «Industry 4.0» revolution in consumer-firm relations
- Starving for insights: consumer social data and Explainable-AI
- Apulian tourism growth: the C-BAS project
  - ➢ Modeling tourists' textual reviews
  - ➢ Setting up a rating-forecast model
  - ➢ What factors drive tourists' choices?
- Conclusions
- Future perspectives

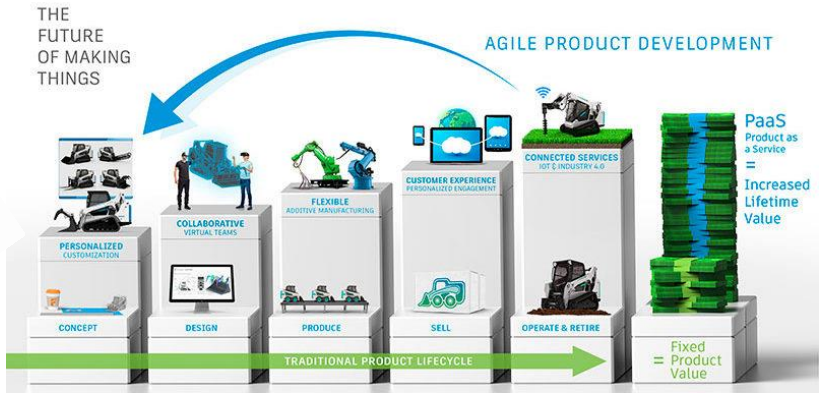
Politecnico di Bari
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

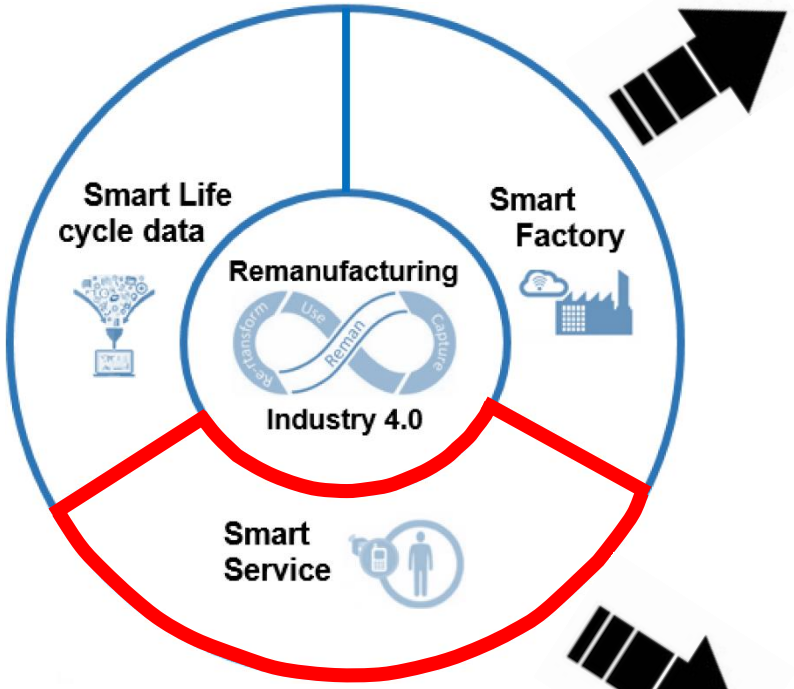
«Industry 4.0» revolution in consumer-firm relations



«Industry 4.0» revolution in consumer-firm relations

Customized Production





Strong involvement

Customers as co-producers

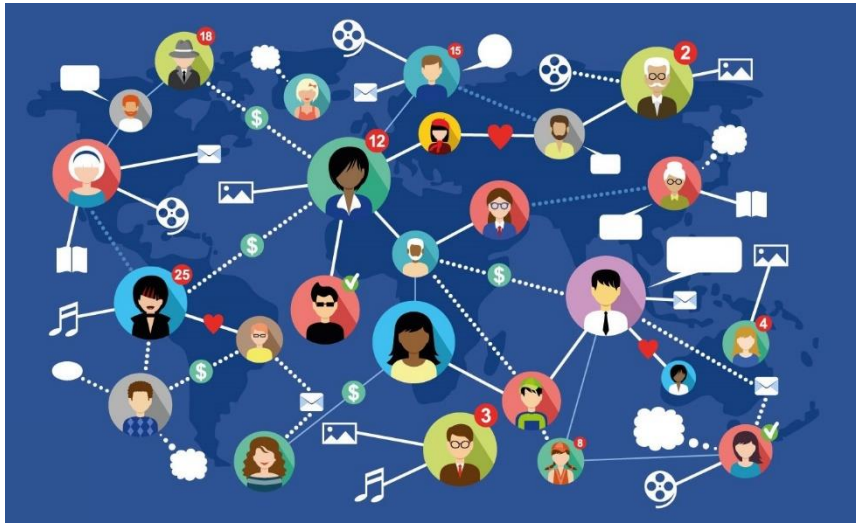
Politecnico di Bari

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Starving for insights: Consumer social data and Explainable-AI

- Consumers can influence each other (e.g.: Online Social Networks, Word-of-Mouth)

Locally…



… and globally!

- Analyzing and forecasting customers' needs and flavours is of crucial importance for every kind of business

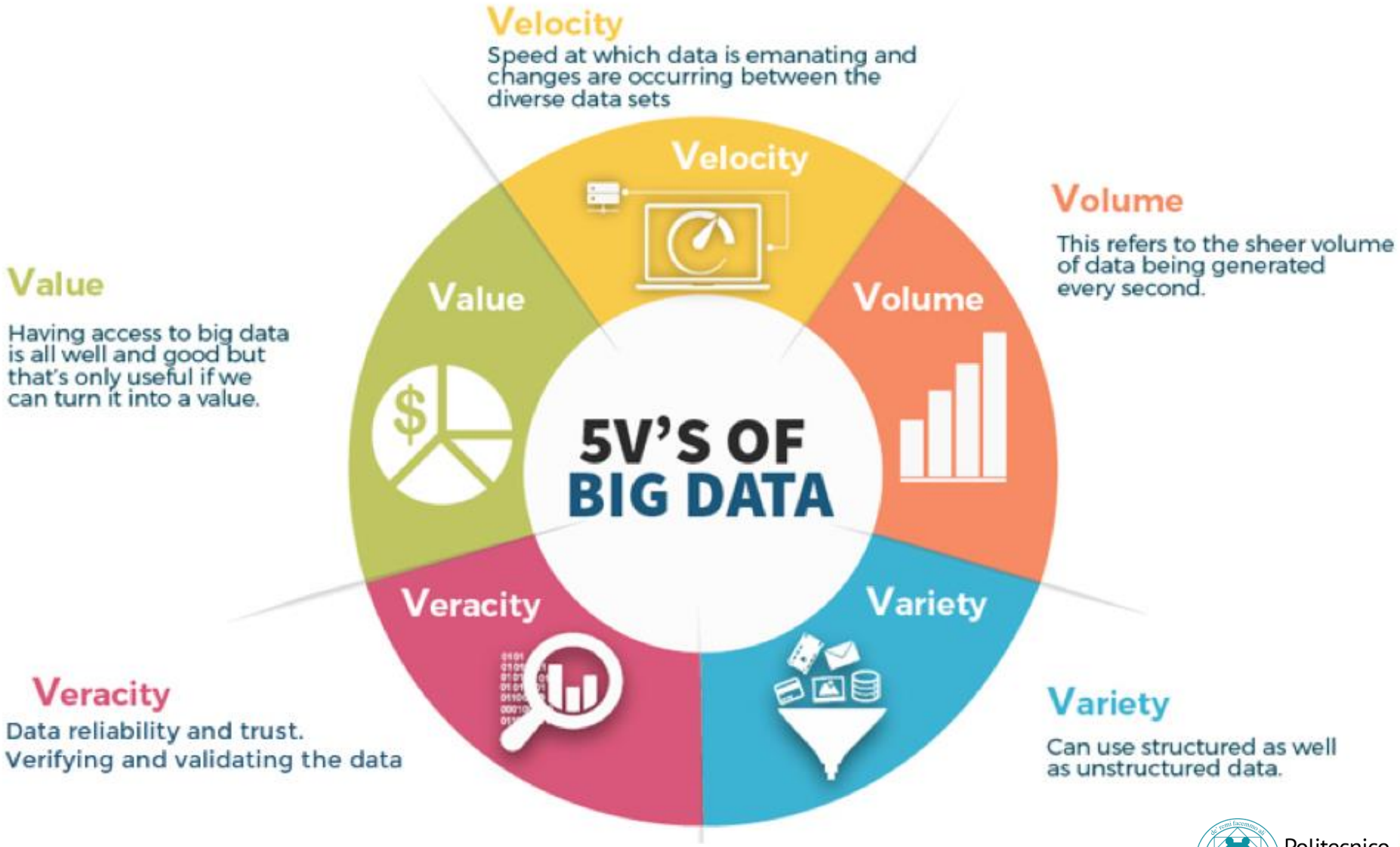
Taming Big Data…

Politecnico di Bari

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

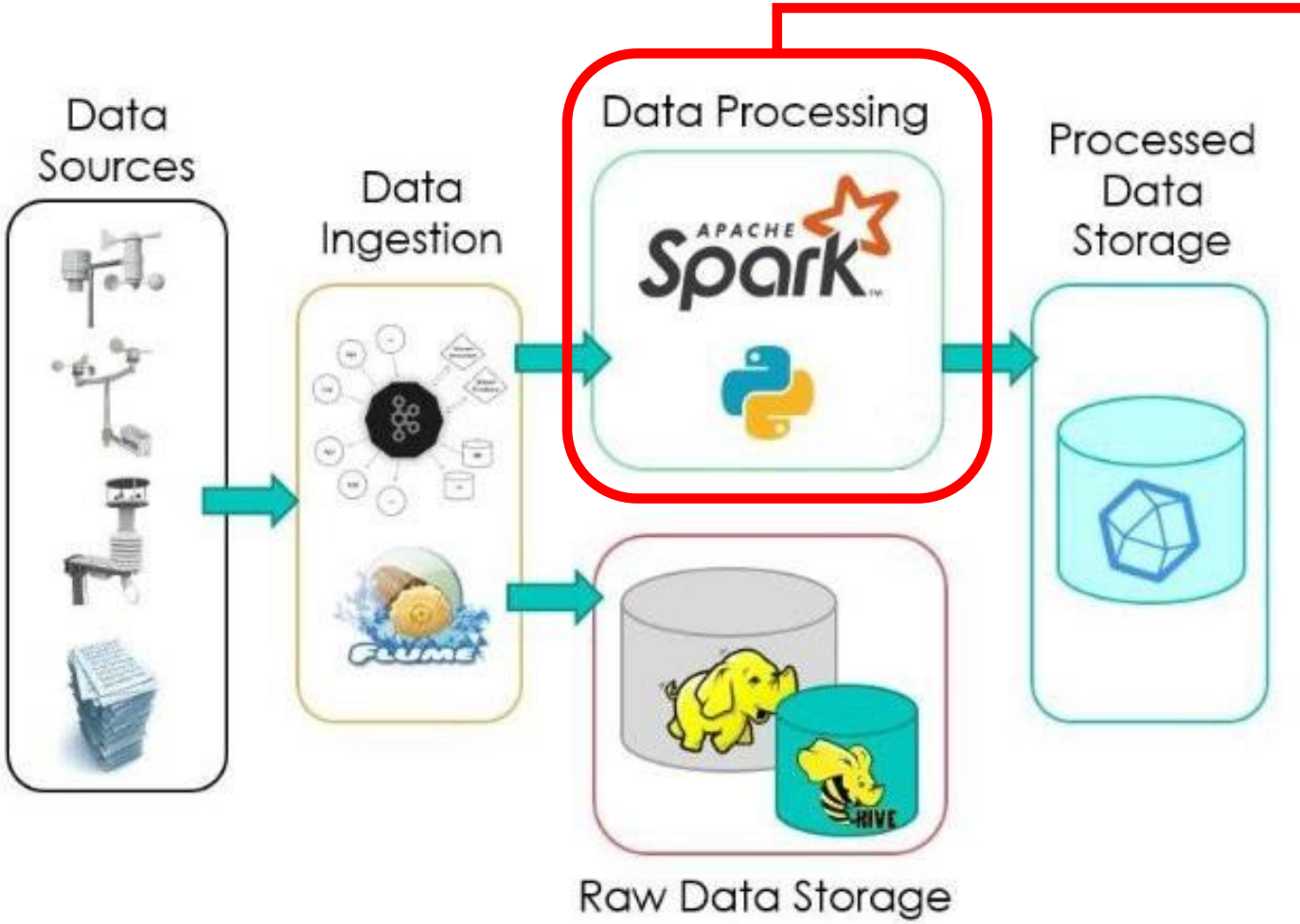
Starving for insights: Consumer social data and Explainable-AI

- A brief overview over big data: definition and characteristics

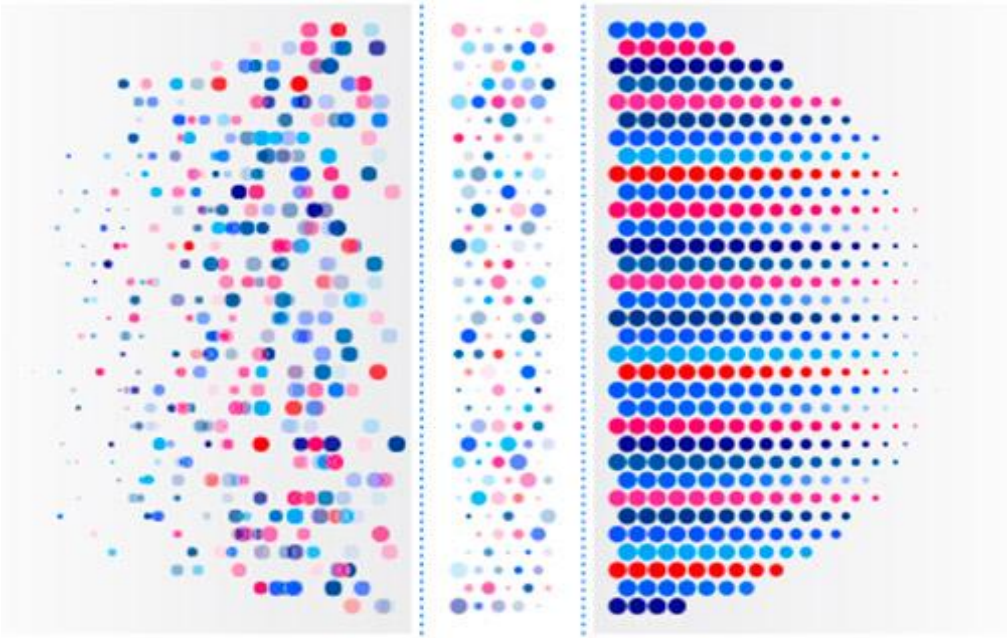


Starving for insights: Consumer social data and Explainable-AI

- A brief overview over big data: storage and analytics



ARTIFICIAL INTELLIGENCE

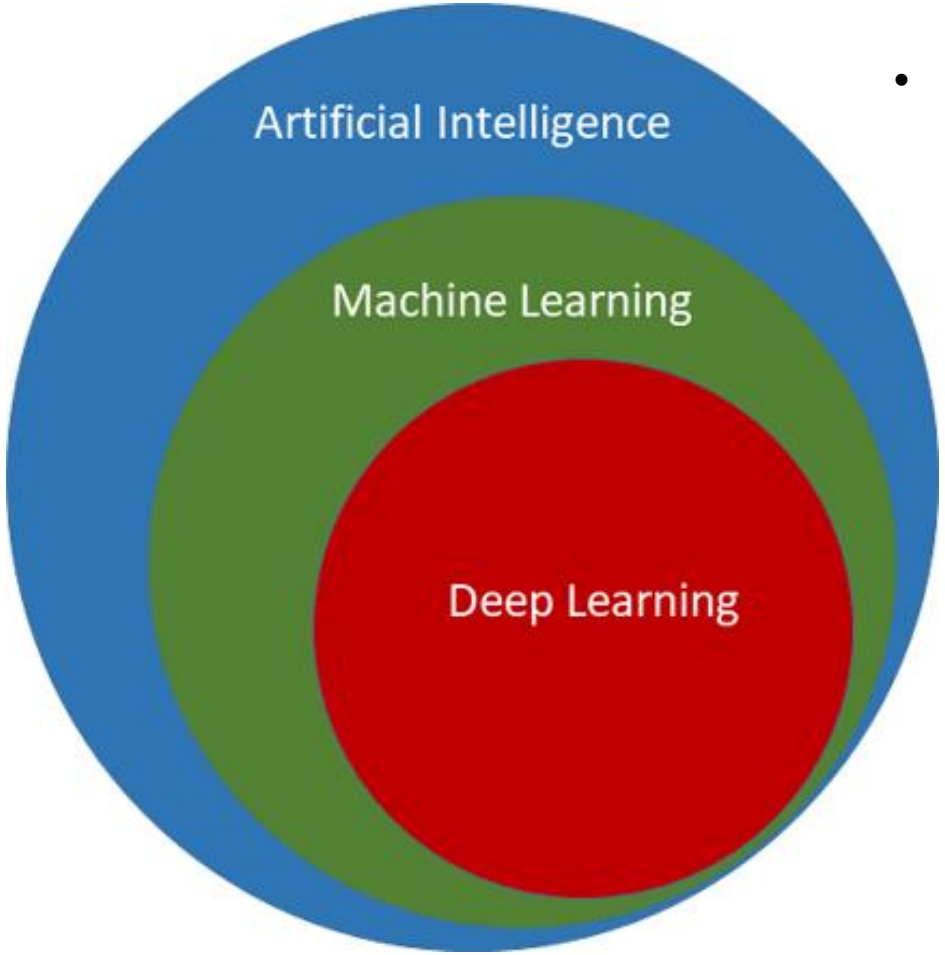


BIG DATA **SOLUTION**

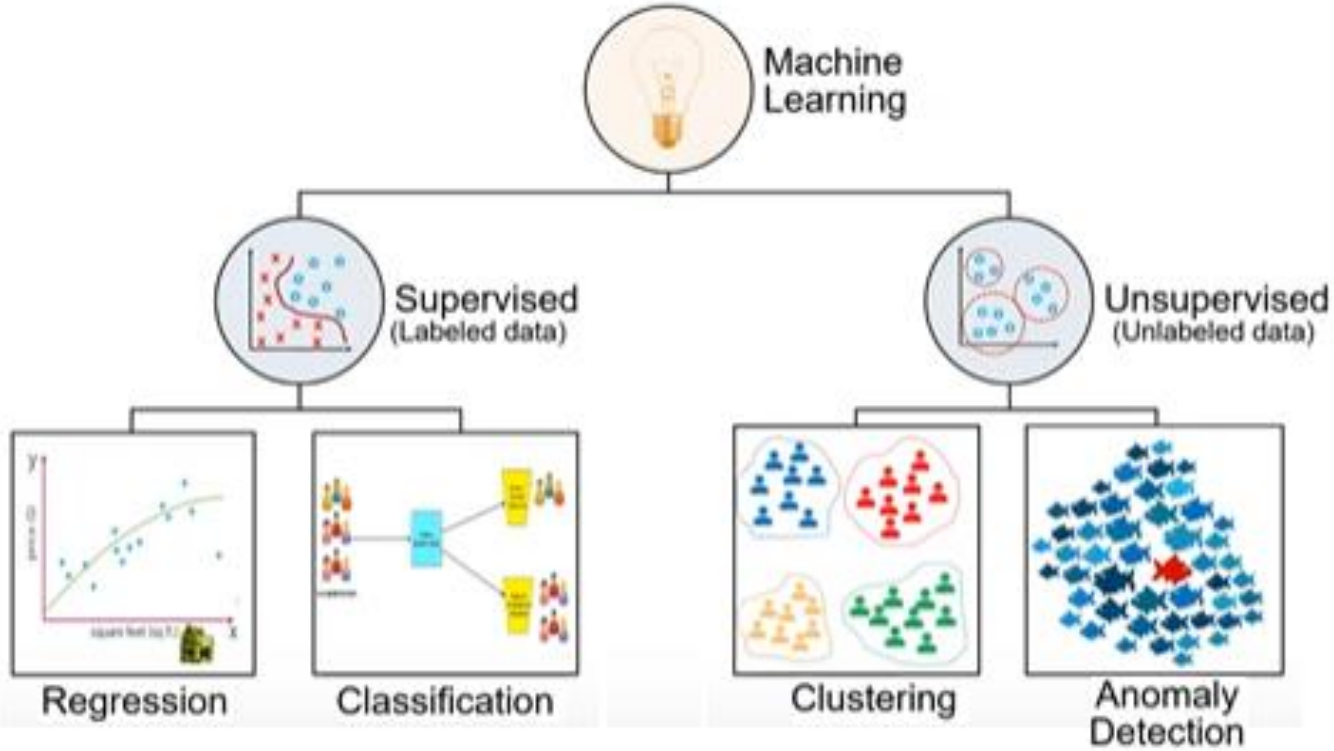
Politecnico di Bari
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Starving for insights: Consumer social data and Explainable-AI

- AI and Machine Learning

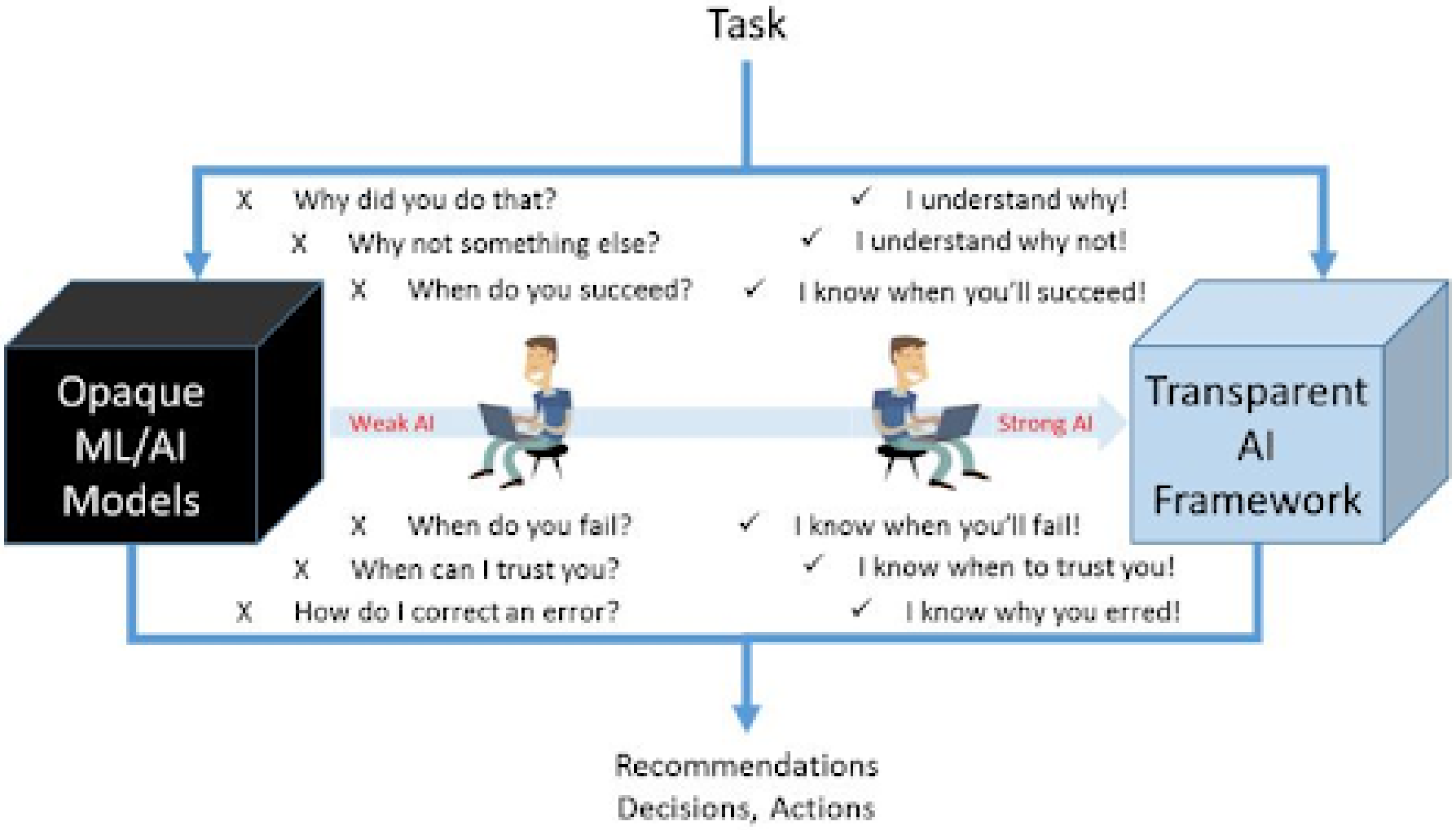


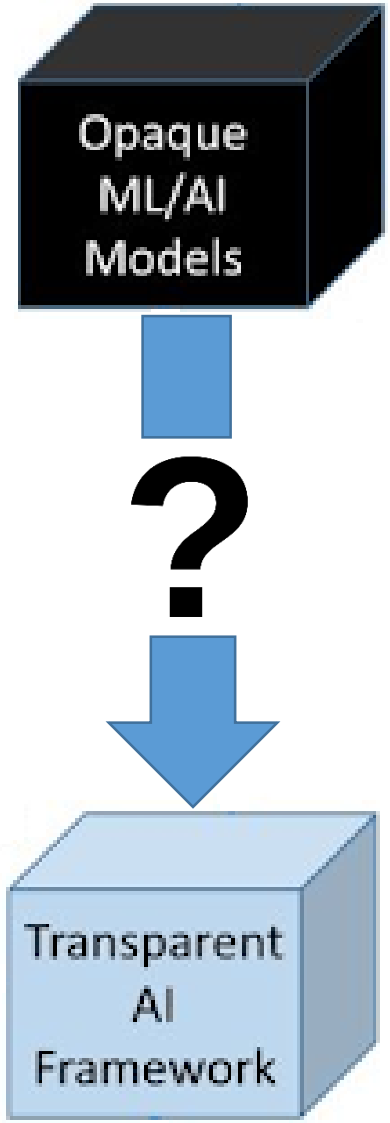
- **Artificial Intelligence**: Solving tasks requiring human intelligence
- **Machine Learning**: subset of AI which allows a machine to automatically learn from past data without programming explicitly.



Starving for insights: Consumer social data and Explainable-AI

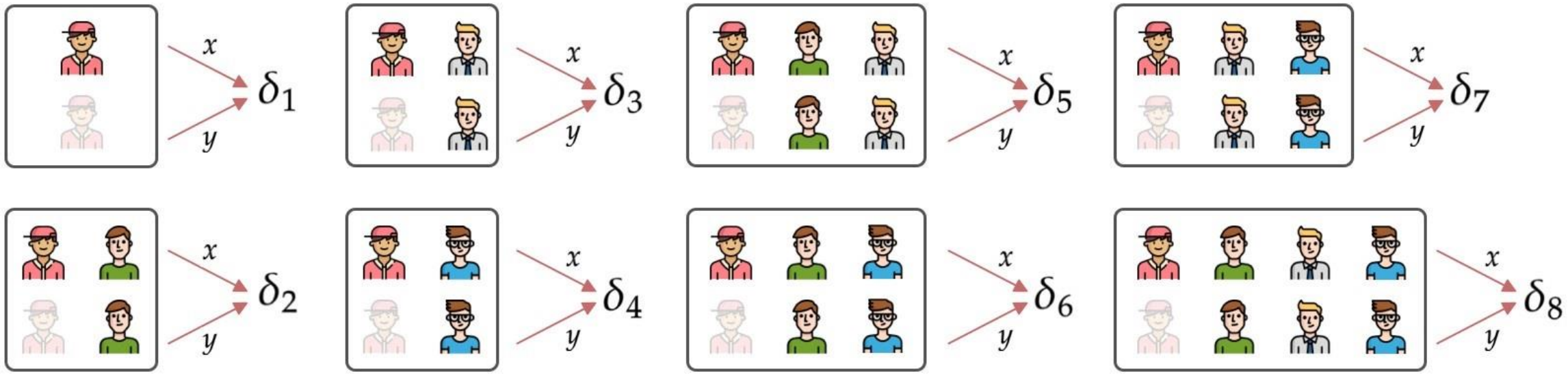
- About Explainability in AI: basics





Starving for insights: Consumer social data and Explainable-AI

- About Explainability in AI: Shapley Values – A tool from cooperative game theory



The Shapley value for member

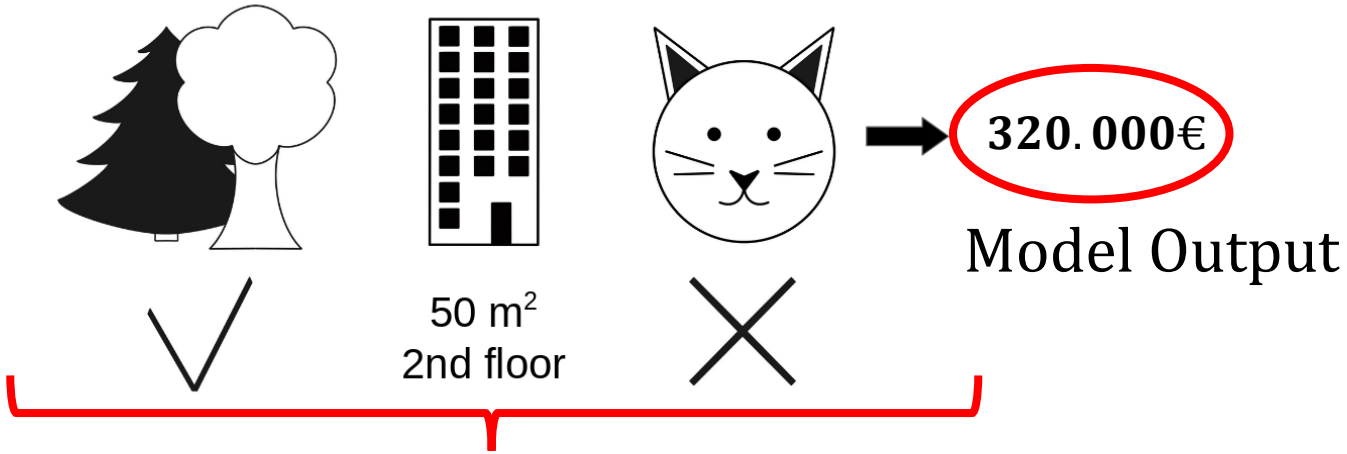
$$\phi_i = \frac{\delta_1 + \delta_3 + \delta_4 + \delta_5 + \delta_6 + \delta_7 + \delta_8}{8}$$

Starving for insights: Consumer social data and Explainable-AI

- About Explainability in AI: Shapley Values in AI

<u>Each feature value is a «player» in a game where the prediction is the «payout» and the «gain» is the difference between model prediction and average prediction</u>

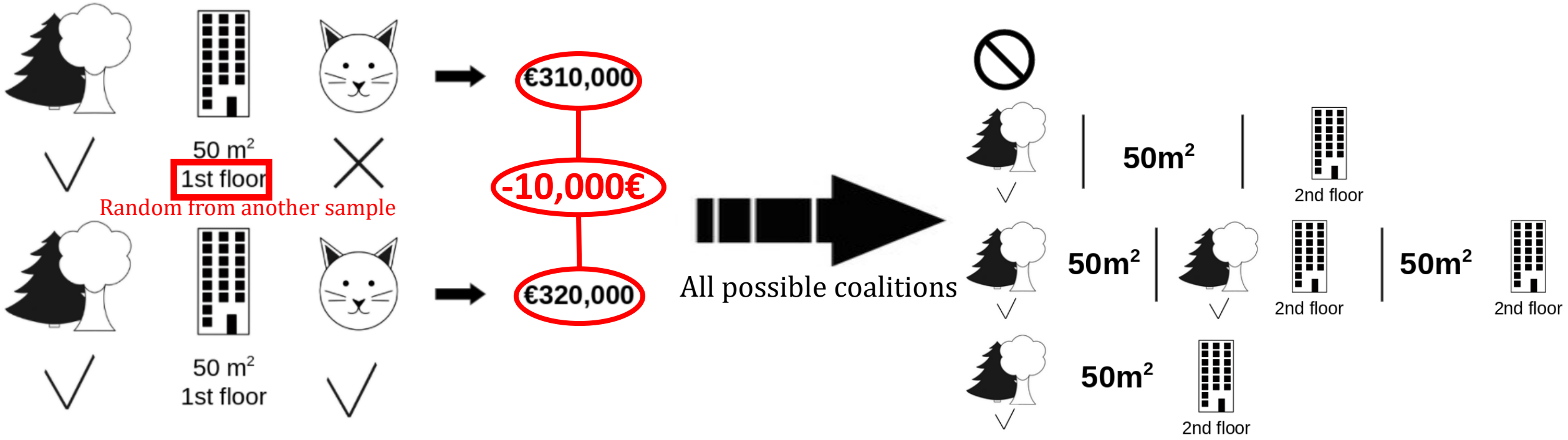
Example: Prediction of house prices based on four features: (1) nearby park; (2) apartment area; (3) apartment floor; (4) cat-banned or not.



<u>Question</u>: How much has the "cat-banned" feature value contributed to the prediction compared to the average prediction?

Starving for insights: Consumer social data and Explainable-AI

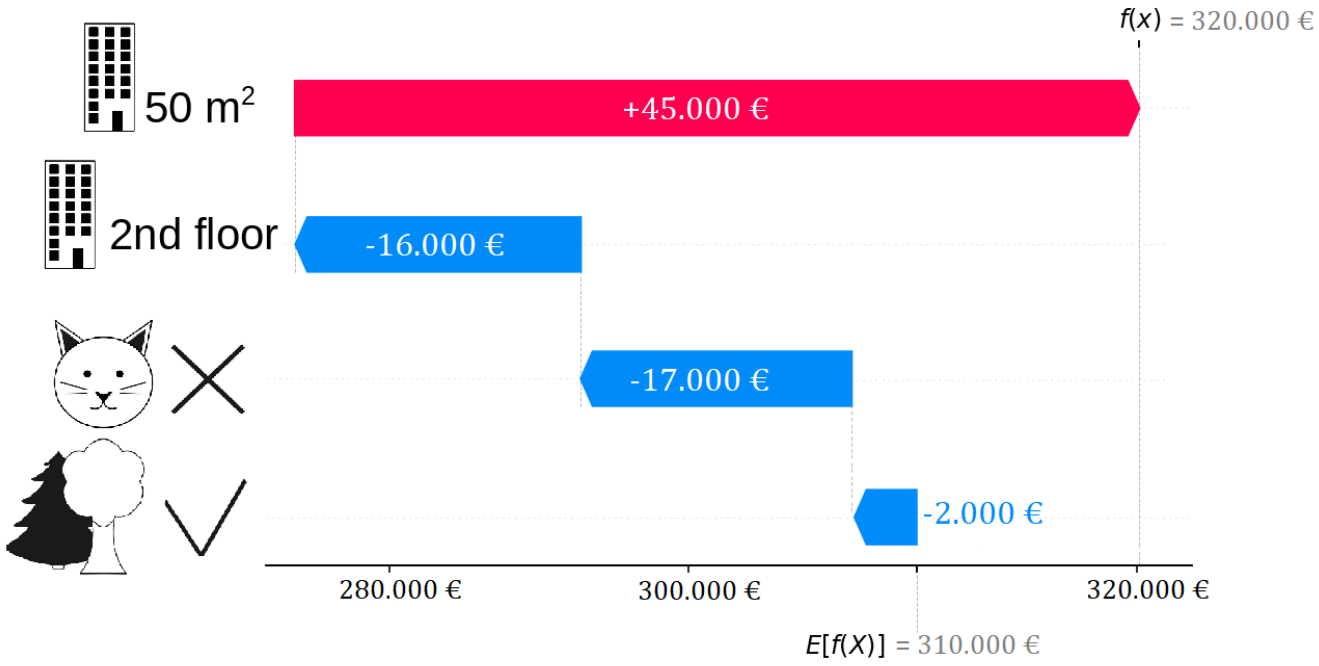
- Following Steps for Shapley values calculation



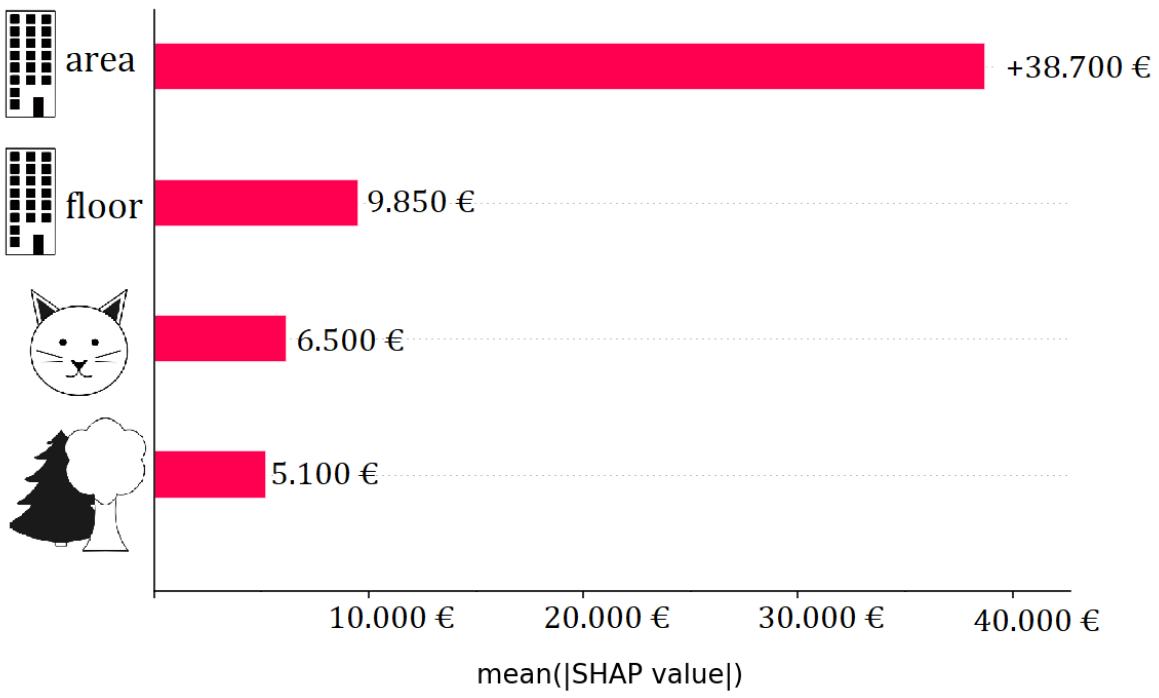
<u>The Shapley value is the average of all the marginal contributions to all possible coalitions.</u>

Starving for insights: Consumer social data and Explainable-AI

- Graphically:



Waterfall plot
(single sample)



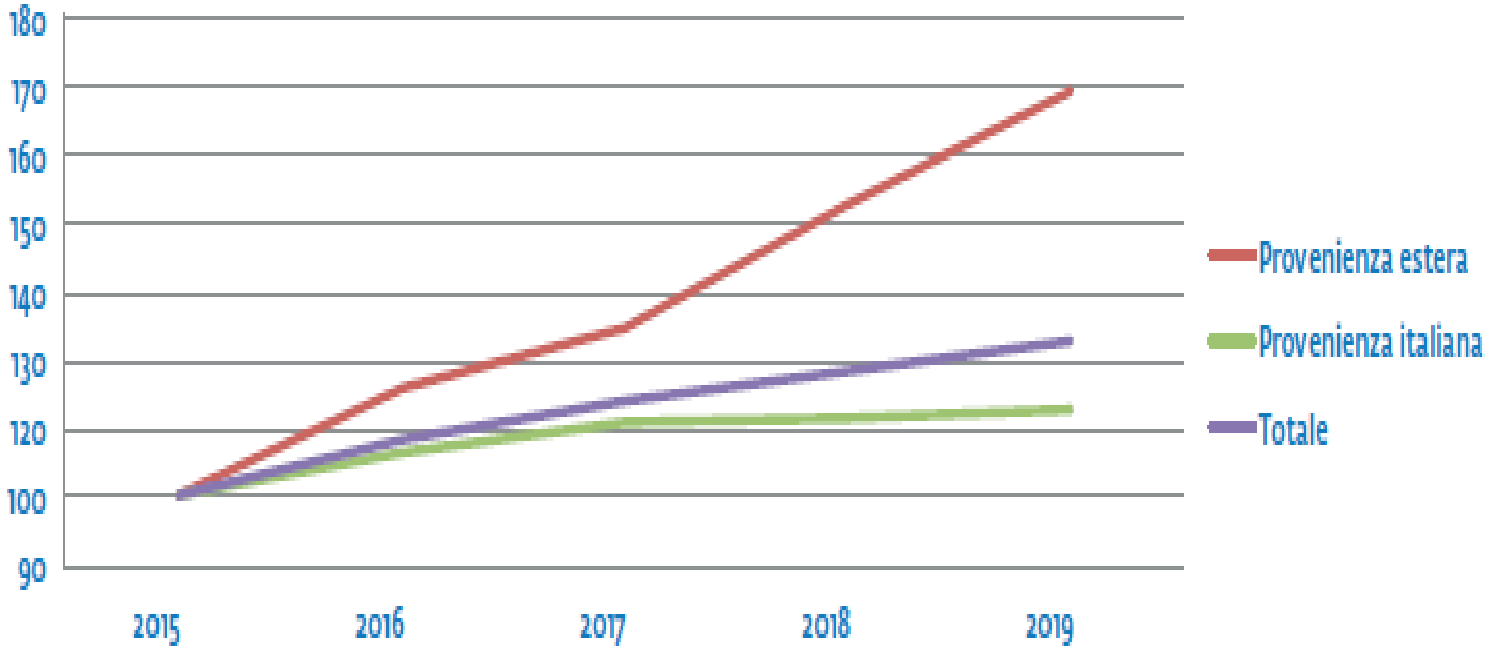
Mean Absolute value of Shapley
(entire dataset)

Apulian tourism growth: the C-BAS project

- «C-BAS» acronym for «Consumer-Behaviour Analysis System»
- «C-BAS» stems from a simple observation…

Apulian tourism went through a massive growth in the past 5 years

… and a simple question: **why?**

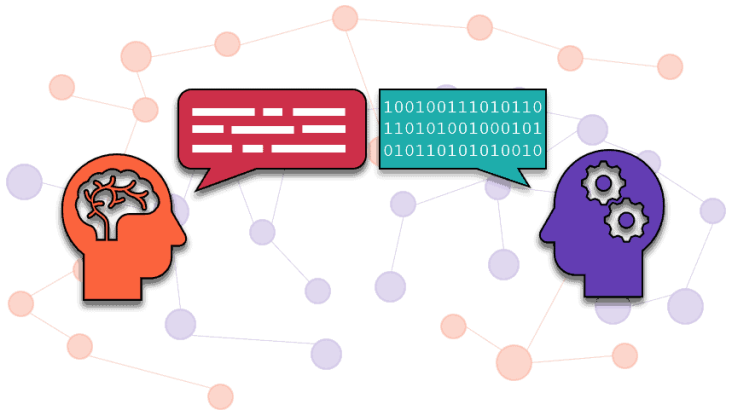


Source: Osservatorio Turistico regionale. ISTAT data, 2019. Index for 2015=100.

Apulian tourism growth: the C-BAS project

- Overall flowchart

Scraping Social Data
Text + Rating



Natural Language Processing (NLP)

Machine Learning
(CLASSIFICATION TASK)

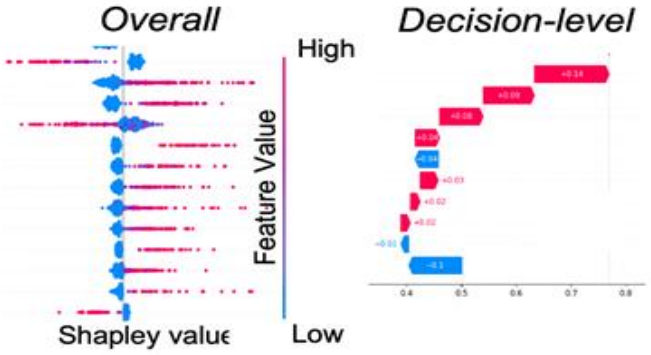
Negative

Rating < 3

Positive

Rating ≥ 3

Explainability



Wow, we visited Monopoli again this year and immediately decided to visit the wonderful Antonella and Gianni, to book 2 bike tours in our week's holiday, after booking with them last year. They both remembered us and immediately made us feel welcome.

Politecnico di Bari
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Modeling tourists' textual reviews

- NLP techniques: Sentiment Filtering and TF-IDF Matrix

Tripadvisor reviews about Apulian facilities (≈ 15.000 reviews; 1.000 facilities)

Sentiment filtering

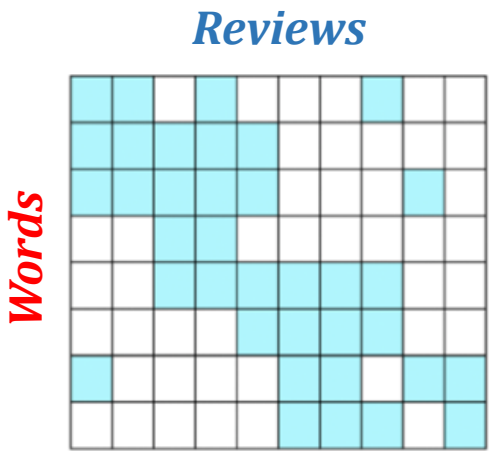
- **Lower case transformation**
- **Removing Stop-Words (e.g. articles, conjunctions)**
- **Stemming (e.g. «fishing», «fished» → «fish»)**

$$T_{i,j} = \underbrace{\boxed{\left(\frac{n_{ij}}{|d_j|}\right)}}_{\textit{Term Frequency}} \times \underbrace{\boxed{\left(\log \frac{|D|}{|\{d: i \in D\}|}\right)}}_{\textit{Inverse Document Frequency}}$$

n_{ij} = number of occurrences of word i in document j

$|d_j|$ = number of words in document j

$|D|$ = number of documents

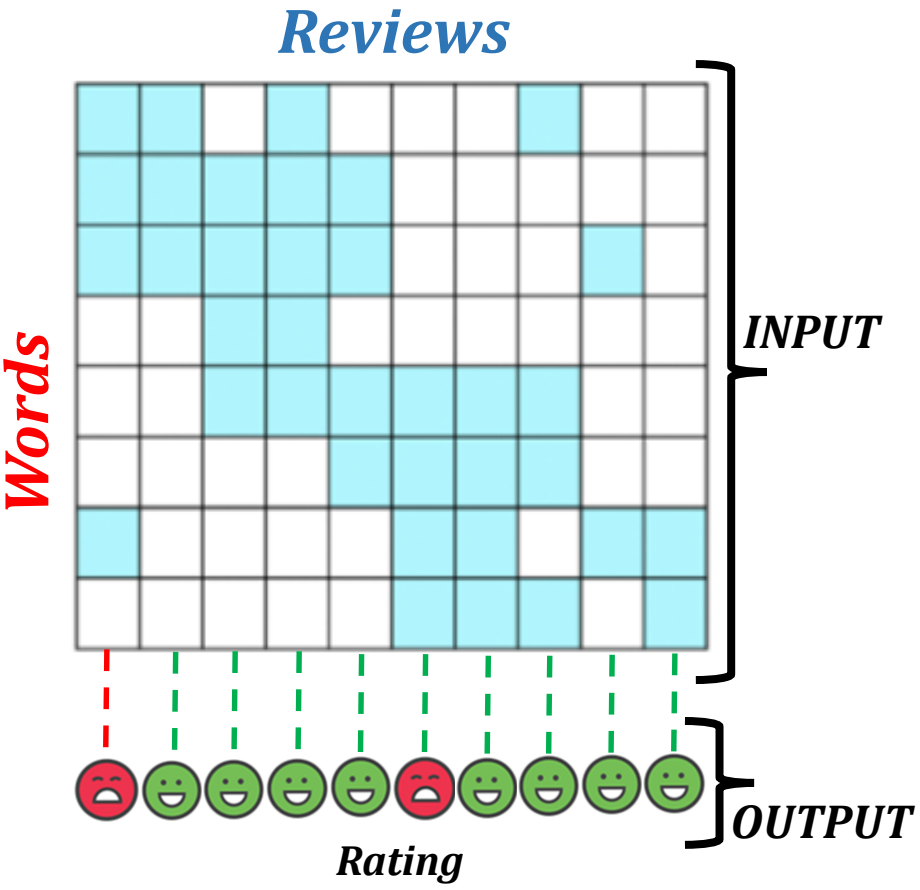


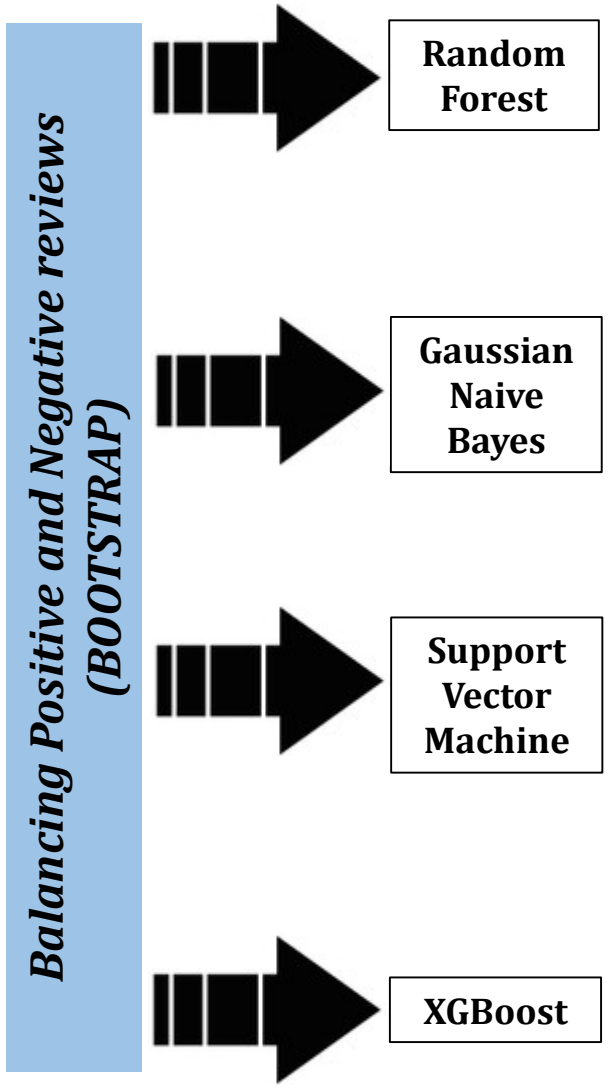
= T

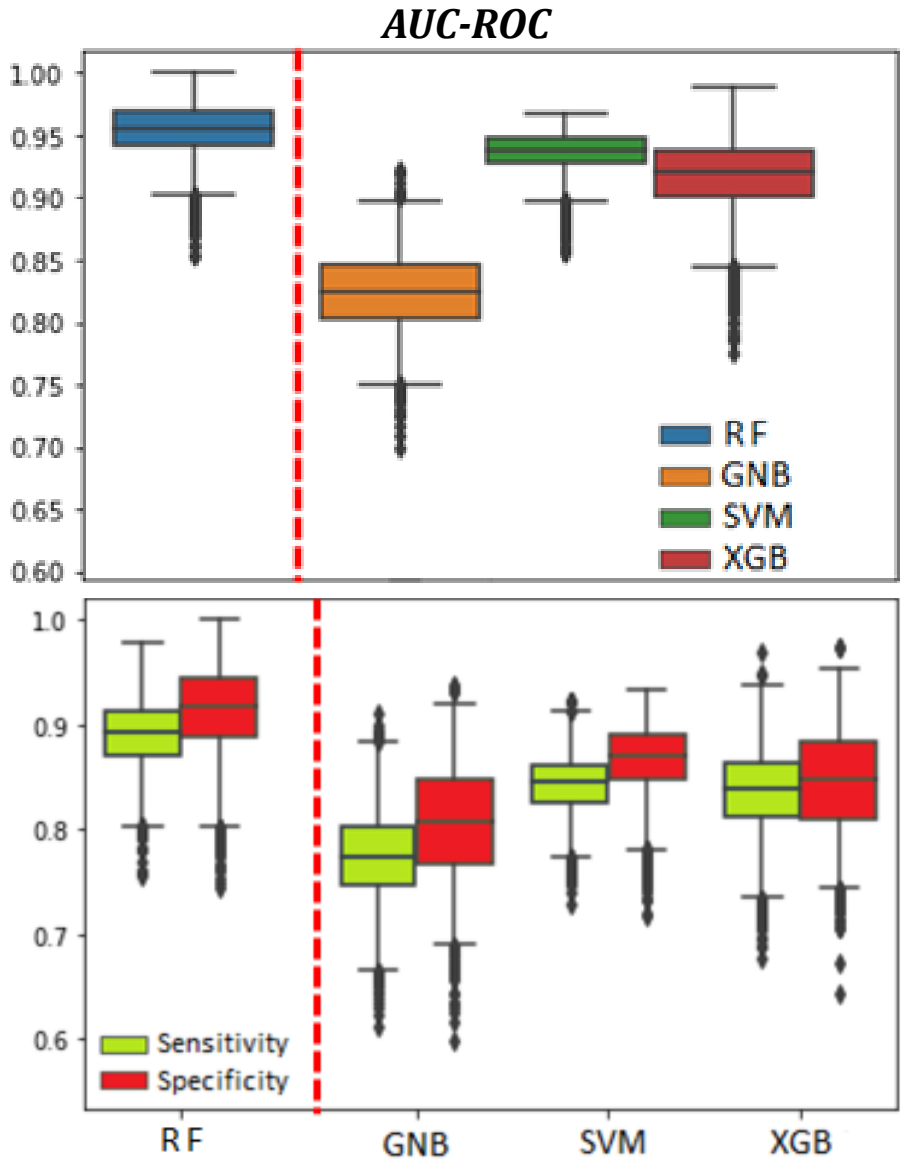
(TF-IDF Matrix)

Setting up a rating-forecast model

- Not just a single model (No-free-lunch theorem)

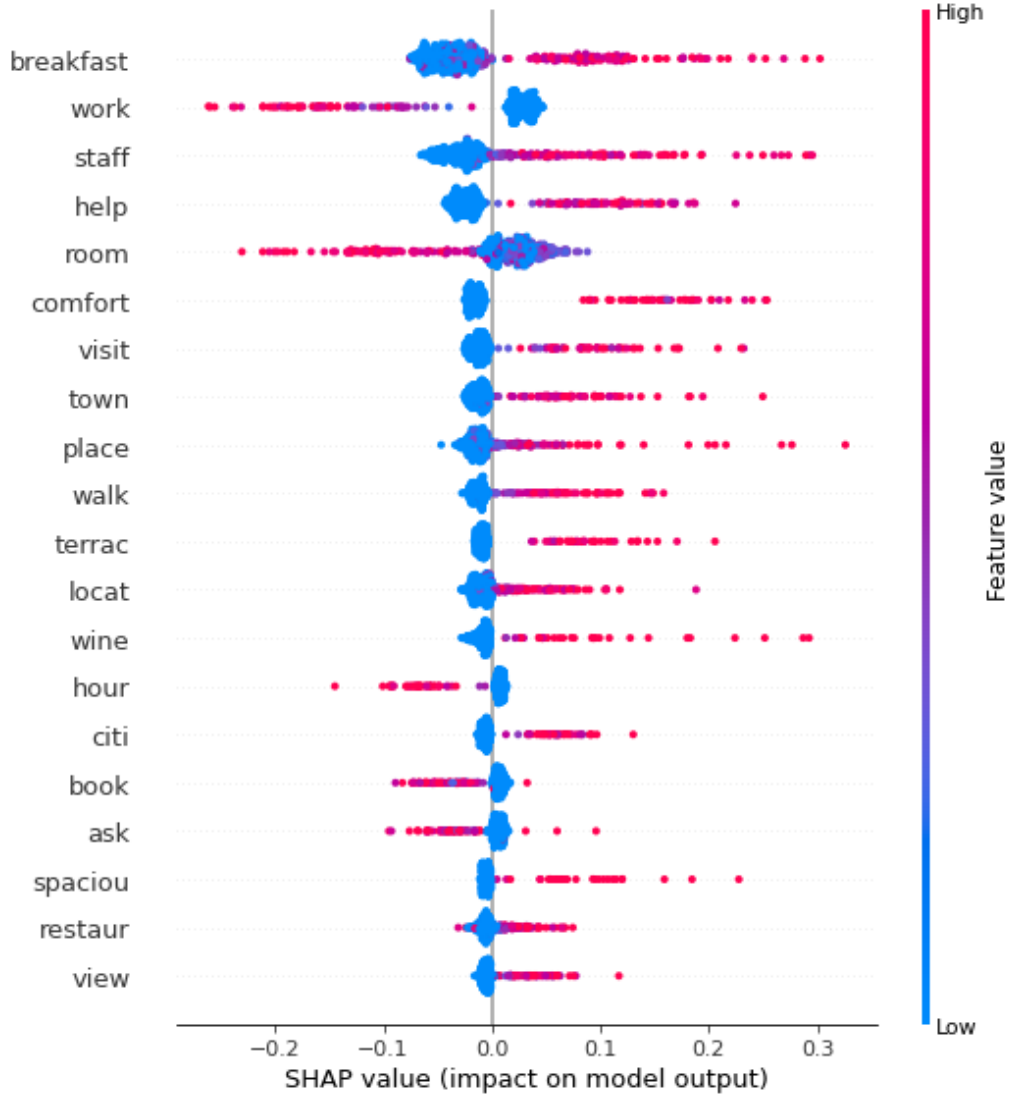




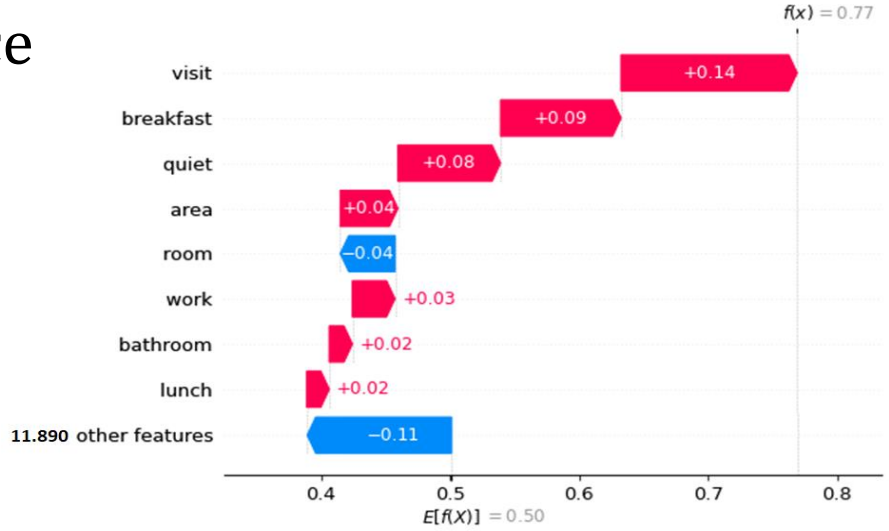


What factors are driving tourism growth?

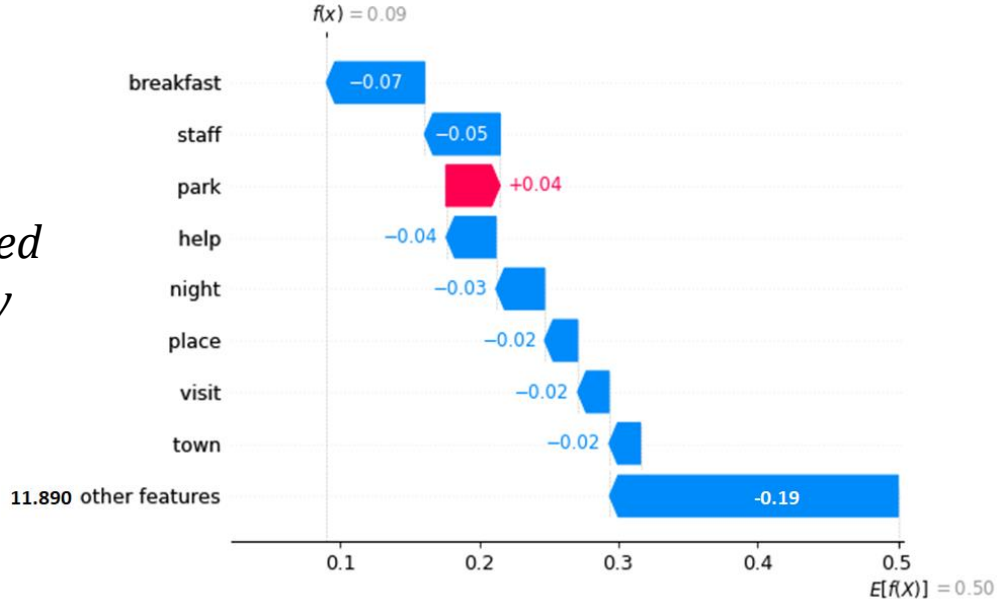
- Shapley values at work for tourim-business intelligence



Correctly classified positive review



Correctly classified negative review



Conclusions

- Words referring to the semantic area of food (**«breakfast», «wine», «restaur»**) **positively impact on ratings**. This is an aspect to take care of in oder to leverage tourism-business.

- **«Staff»**, **«help»** and «**(old) town**» **positively influence tourist experiences and ratings**. Staff quality is an aspect that tourists consider to evaluate their experience: it should not be overlooked. Moreover, livability of old town must be a focus of both private and public stakeholders.

- **«Room»**, on the contrary, **negatively affects ratings**: the room-quality and comfort is an aspect on which tourism businessses need to focus their efforts.

Politecnico di Bari
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Future perspectives

- Extracting reviews from other online social networks (e.g.: Booking, Twitter, Facebook)

- Adding tourist information (e.g. age, marital-status, travel type) to determine «communities of tourists» with same habits and create a recommendation system.

- Applying Deep Learning models to process reviews: LSTM and GRU Recurrent Neural Networks (wok in progress)

THANK YOU !

For more info, please mail me to: francesco.denicolo@poliba.it

Politecnico di Bari
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

BACKGROUND MATERIAL

Sentiment filtering

- Sometimes rating does not express what is written in the reviews:

"We were lucky to get the apartment in mon amour b &b which has a huge bedroom and a large balcony with a view of the sea. The location couldn't be better you have the great views but although it's only a short walk to the beach and town it's lovely and quiet. There is a small galley kitchen and a very small shower. would love to go back!"

SENTIMENT SCORE
(between -1 and +1)

0,96

RATING
(between 1 and 5 stars)

1 star

BINARY SENTIMENT SCORE
(«0» if sentiment score ≤ 0; «1» if sentiment score > 0)

BINARY RATING
(«0» if rating ≤ 3; «1» if rating > 3)

1 — *MISMATCH* — 0

Sentiment filtering

- Sometimes rating does not express what is written in the reviews:

"Stayed here recently for three nights. We filled in the gp coordinates and arrived in the driveway of a family home about 50 m from the main entrance gates--quite eventful--then when we arrived at the main gate nobody answered for at least 10 minutes. Our apartment style rooms were good and spacious. Pool was lovely. The breakfast was poor in my eyes. Choice of fruits was limited and cereals as well. I decided to have cornflakes one morning but there was no cold milk in a jug to pour over them--I asked for some and they took my plate of cornflakes away and drowned them in milk---it spoilt them and a didn't eat them. Also breads limited--plenty of cakes though if you like that sort of thing."

SENTIMENT SCORE
(between -1 and +1)

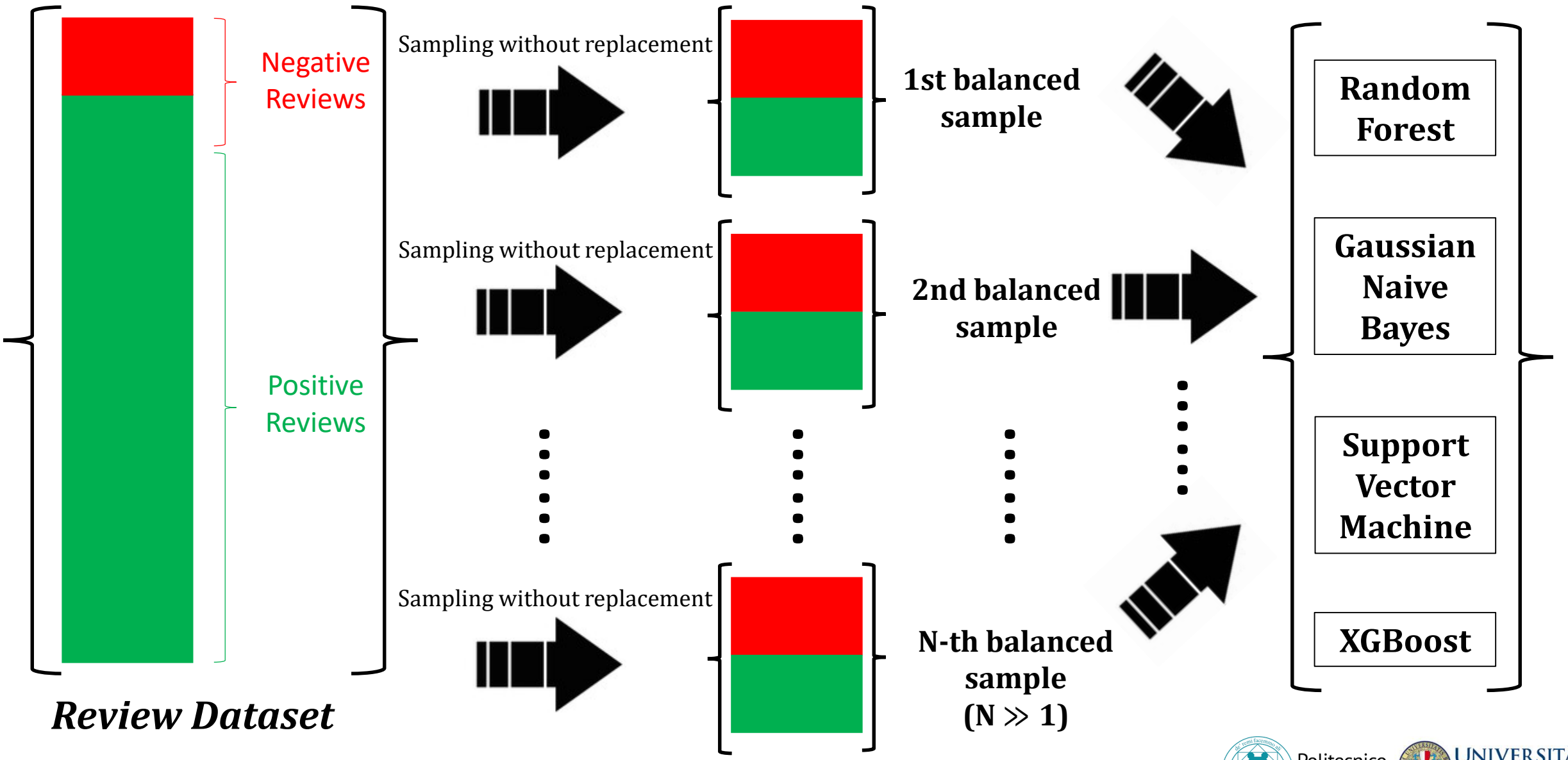
$-0{,}54$

RATING
(between 1 and 5 stars)

4 stars

BINARY SENTIMENT SCORE
(«0» if sentiment score ≤ 0; «1» if sentiment score > 0)

BINARY RATING
(«0» if rating ≤ 3; «1» if rating > 3)

0 — *MISMATCH* — 1

Bootstrap for balancing positive and negative reviews