EXPLAINING THE BLACK BOX: WHEN LAW CONTROLS AI

Presentation

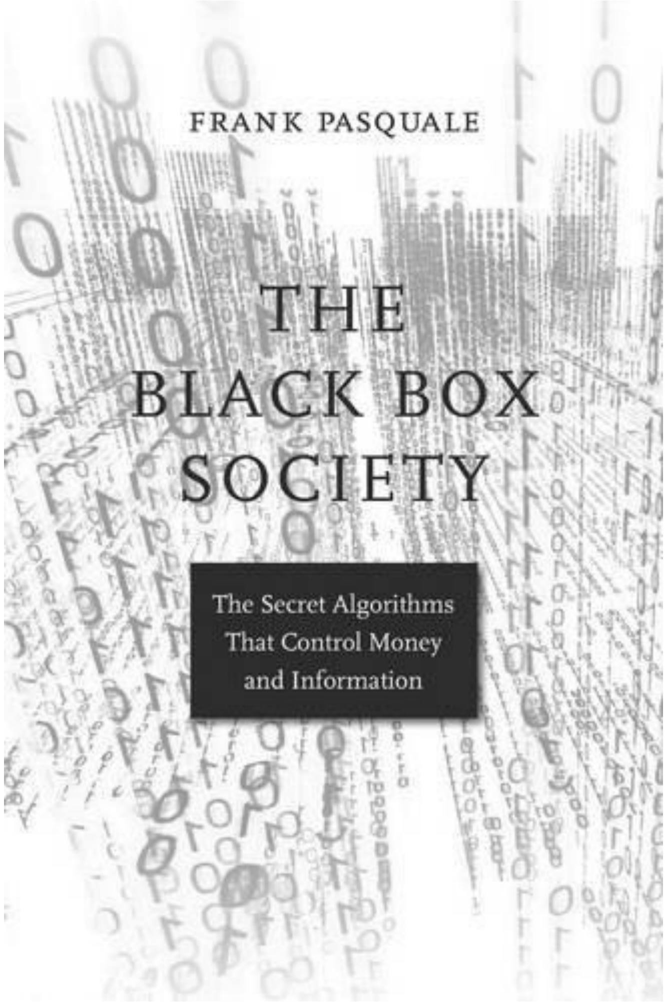
A. Bibal, M. Lognoul, A. de Streel and B. Frenay,
“Legal Requirements on Explainability in Machine Learning”,
Artificial Intelligence and Law, 2020, forthcoming

Alexandre de Streel

Joint Academic Director, CERRE
Professor of EU Law, University of Namur

Benoît Frenay

Associate Professor,
University of Namur

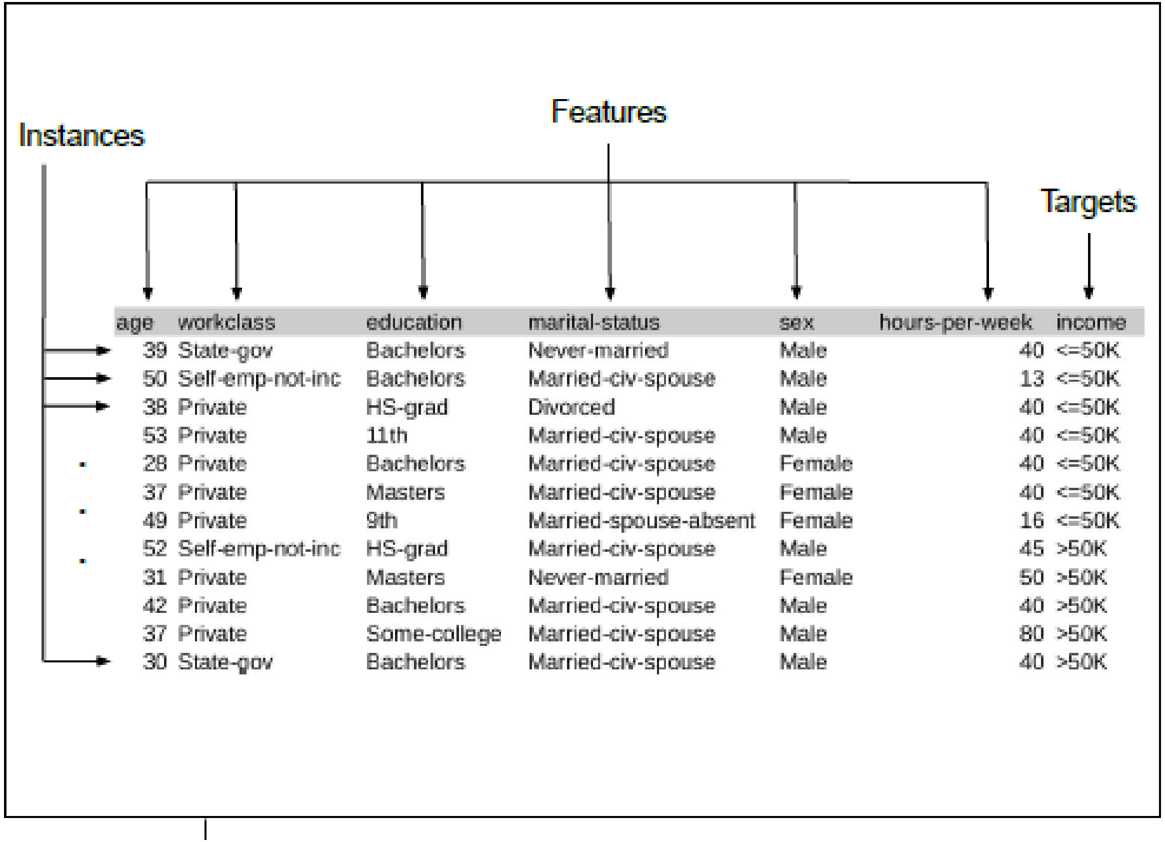


FRANK PASQUALE

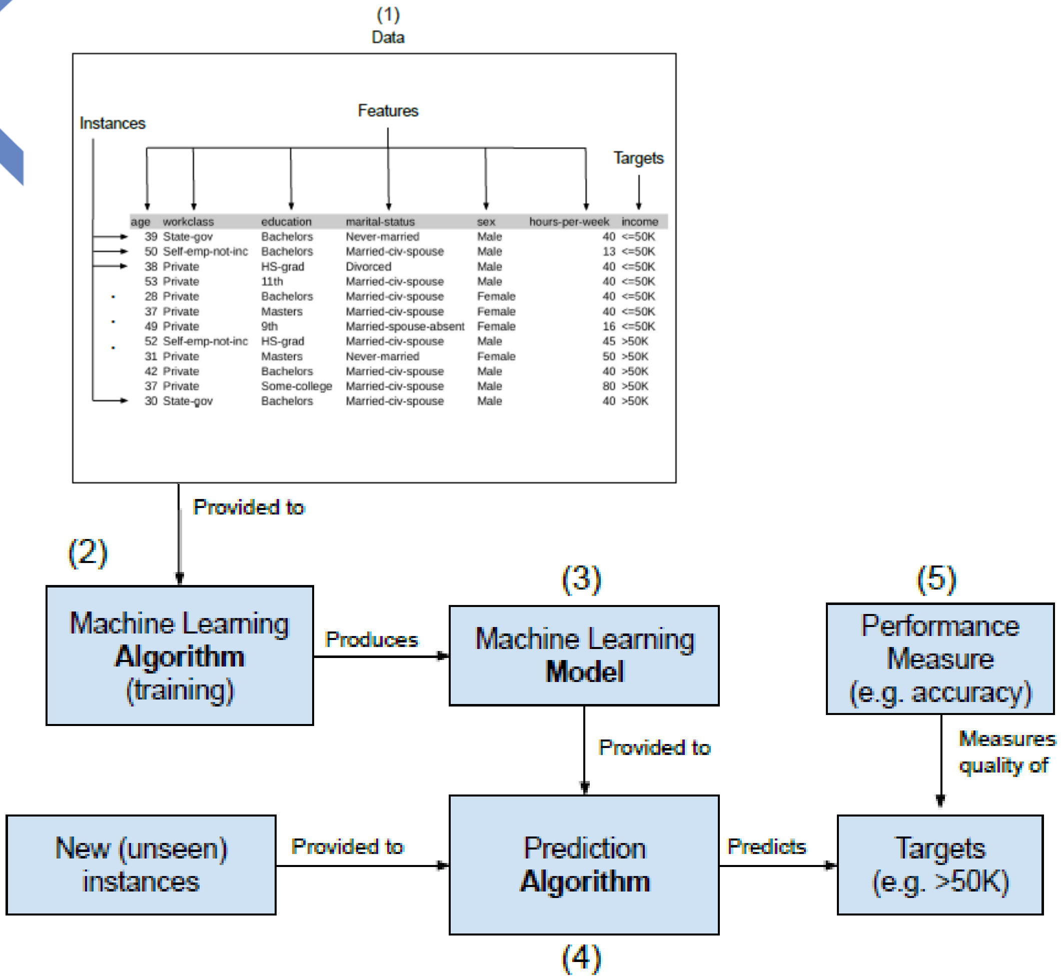
THE BLACK BOX SOCIETY

The Secret Algorithms That Control Money and Information

(1)
Data


Instances
Features
Targets

age	workclass	education	marital-status	sex	hours-per-week	income
39	State-gov	Bachelors	Never-married	Male	40	<=50K
50	Self-emp-not-inc	Bachelors	Married-civ-spouse	Male	13	<=50K
38	Private	HS-grad	Divorced	Male	40	<=50K
53	Private	11th	Married-civ-spouse	Male	40	<=50K
28	Private	Bachelors	Married-civ-spouse	Female	40	<=50K
37	Private	Masters	Married-civ-spouse	Female	40	<=50K
49	Private	9th	Married-spouse-absent	Female	16	<=50K
52	Self-emp-not-inc	HS-grad	Married-civ-spouse	Male	45	>50K
31	Private	Masters	Never-married	Female	50	>50K
42	Private	Bachelors	Married-civ-spouse	Male	40	>50K
37	Private	Some-college	Married-civ-spouse	Male	80	>50K
30	State-gov	Bachelors	Married-civ-spouse	Male	40	>50K



(1)
Data

Instances

Features

Targets

age	workclass	education	marital-status	sex	hours-per-week	income
39	State-gov	Bachelors	Never-married	Male	40	<=50K
50	Self-emp-not-inc	Bachelors	Married-civ-spouse	Male	13	<=50K
38	Private	HS-grad	Divorced	Male	40	<=50K
53	Private	11th	Married-civ-spouse	Male	40	<=50K
28	Private	Bachelors	Married-civ-spouse	Female	40	<=50K
37	Private	Masters	Married-civ-spouse	Female	40	<=50K
49	Private	9th	Married-spouse-absent	Female	16	<=50K
52	Self-emp-not-inc	HS-grad	Married-civ-spouse	Male	45	>50K
31	Private	Masters	Never-married	Female	50	>50K
42	Private	Bachelors	Married-civ-spouse	Male	40	>50K
37	Private	Some-college	Married-civ-spouse	Male	80	>50K
30	State-gov	Bachelors	Married-civ-spouse	Male	40	>50K

Provided to

(2)
Machine Learning **Algorithm** (training)

Produces

(3)
Machine Learning **Model**

Provided to

(5)
Performance Measure (e.g. accuracy)

Measures quality of

New (unseen) instances

Provided to

Prediction **Algorithm**

(4)

Predicts

Targets (e.g. >50K)

Meaning of Explainability
High Level Group on AI

Explicability is crucial for building and maintaining users’ trust in AI systems.

- **Processes need to be transparent**
- **Capabilities & purpose of AI systems need to be openly communicated**
- **Decisions need to be – to the extent possible – explainable** to those directly and indirectly affected.

Without such information, a decision cannot be duly contested.

Meaning of Explainability
High Level Group on AI

Explanation as to **why** a model has generated a particular output or decision (and **what** combination of input factors contributed) is not always possible.

‘BLACK BOX’ ALGORITHMS REQUIRE SPECIAL ATTENTION

Other explicability measures may be required, provided that the system as a whole respects fundamental rights:

- **traceability**
- **auditability**
- **transparent communication on system capabilities**

Meaning of Explainability
High Level Group on AI

The degree to which explicability is needed is

highly dependent on the context & severity of the consequences

if that output is erroneous or otherwise inaccurate.

Meaning of Explainability
High Level Group on AI

EXPLAINABILITY: ability to explain the **technical processes** of an AI system & the related human **decisions**.

Technical explainability requires that the decisions made by an AI system can be **understood and traced** by human beings.

Trade-offs might have to be made:

- enhancing a system's explainability - may reduce accuracy
- increasing its accuracy - at the cost of explainability

Meaning of Explainability
High Level Group on AI

If AI system has **significant impact on people’s lives** -> possibility to request suitable explanation of the system’s decision-making process.

Explanation should be timely and adapted to the expertise of the stakeholder concerned (e.g. layperson, regulator, researcher).

Explanations should be available on the degree to which an AI system influences and shapes:

- the organisational decision-making process
- design choices of the system
- the rationale for deploying it

→ business model transparency

Meaning of Explainability
ICO Draft Guidance

- Rationale explanation
- Responsibility explanation
- Data explanation
- Fairness explanation
- Safety & performance explanation
- Impact explanation

Meaning of Explainability
Computer science

- Ability for an abstract **mathematical model** to be **understood** by its users

- **Interpretable models**
  - Simple mathematical expression (e.g. linear models)
  - Representation allows users to understand their mathematical expression (e.g. decision trees)

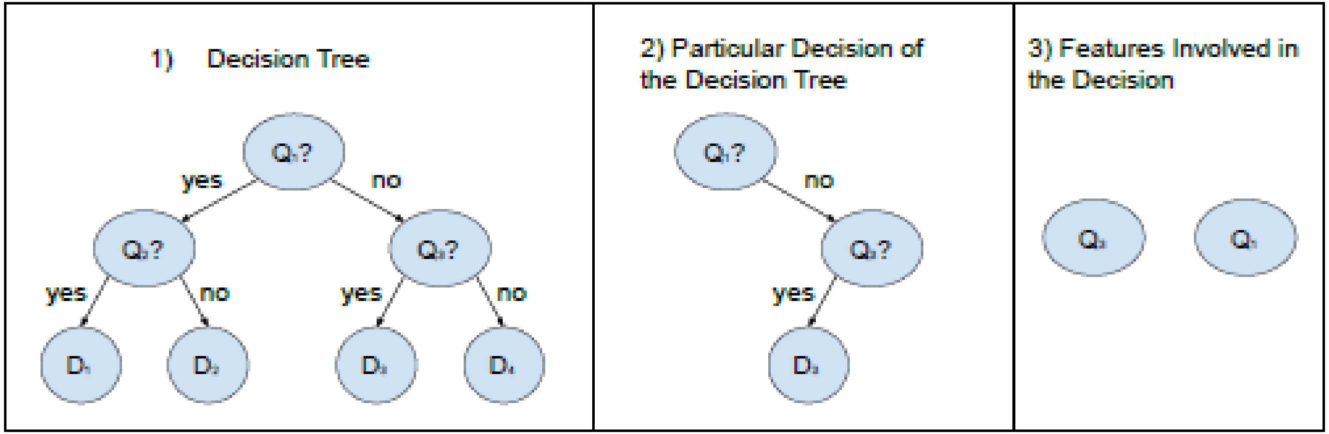
- **Black-box models**
  - Neural networks

Legal Obligations on XAI

	Horizontal legislation	Sector-specific legislation
AI-specific obligations	- Personal data protection: GDPR, Convention 108+ - Consumer acquis - P2B Regulation	- Finance - Health - Automotive …
General obligations	Consumer acquis	

Legal Obligations on XAI

Main features	Directive 2011/83 on Consumer Rights, art. 6(a): *obligation to provide “**the main parameters**" and “**the relative importance of those parameters**"* Regulation 2019/1150 on promoting fairness and transparency of online intermediation services, art. 5: *obligation to provide “the **main parameters**" and “the **relative importance** of those parameters"*
All features	GDPR, art.22 and Guidelines on Automated individual decision-making & Profiling *obligation to provide “the **criteria relied on in reaching the decision**"*
Combination of features	GDPR, art.22 and Guidelines on Automated individual decision-making and Profiling *obligation to provide “the **rationale behind the decision**"*
Whole model	Directive 2014/65 on Markets in Financial Instruments, art. 17 obligation to provide “**information […] about its algorithmic trading and the systems used for that trading**"



1) Decision Tree

2) Particular Decision of the Decision Tree

3) Features Involved in the Decision

Implementation in ML models

Main features	Well developed in ML Linear models: weight to features, strongly and weakly relevant features Black-box models: features sorted by importance
All features	Possible for all ML models Some models show trade-off between accuracy and complexity
Combination of features	Require the use of transparent models: decision tree, linear Create new ones to explain black box models: local explanation
Whole model	Possible only for some models

Issues

50 SHADES OF TRANSPARENCY

Explanation – audit
Types of explanations
Timing: ex ante, ex post

MANY RULES EXIST ALREADY

Non AI specific and AI specific
Horizontal and sector specific

RISK-BASED AND PROPORTIONALITY

TRADE-OFFS

Accuracy & explainability
Rights of users
Rights of AI developers/owners

cerre

Centre on Regulation in Europe

Avenue Louise, 475 (box 10)
1050 Brussels, Belgium

+32 2 230 83 60

info@cerre.eu

cerre.eu

@CERRE_ThinkTank

cerre.eu

Improving network and digital industries regulation