Meditations on First Deployment

A Practical Guide to Responsible Development

EuroPython 2020

Alejandro Saucedo

@AxSaucedo

@AxSaucedo

Hello, my name is Alejandro

@AxSaucedo

Alejandro Saucedo
@AxSaucedo

Engineering Director
Seldon Technologies

Chief Scientist
The Institute for Ethical AI & ML

Governing Council Member-at-Large
Association for Computing Machinery

@AxSaucedo

The magic of programming

You can wake up with an idea
and have a prototype by the end of day/weekend.

Software is eating the world

The future wonders of the world
will be running Python

@AxSaucedo

Critical infrastructure increasing depends on running software

...and regardless of the software / hardware abstractions, the impact will always be human, at an individual and societal level

@AxSaucedo

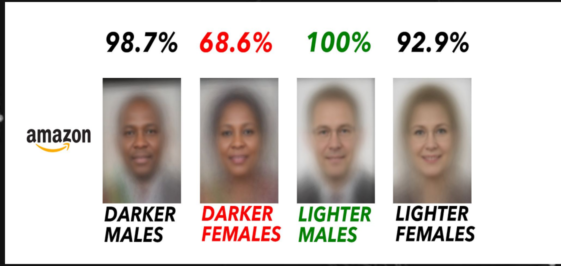
Urgency ~~vs~~ Best Practice AND

@AxSaucedo

Cybersecurity Attacks

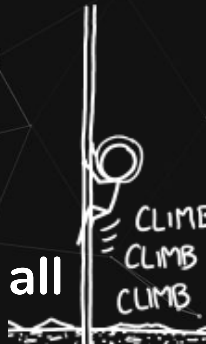
Software Outages

Misuse of personal data



Algorithmic Bias

The impact of a bad solution can be worse than no solution at all



@AxSaucedo

Responsibility Infrastructure

Department / Organisation

- High level Principles
- Governing structure
- Aligned objectives
- Escalation structure

Team / Delivery Process

- Cross functional skillset
- Key domain experts
- Accountability structure
- Principled alignment
- Relevant delivery structure

Individual Practitioner

- Technology best practices
- Most relevant tools
- Competence in field
- Professional responsibility

@AxSaucedo

Professional Responsibility

As software developers we have a growing professional responsibility to our craft

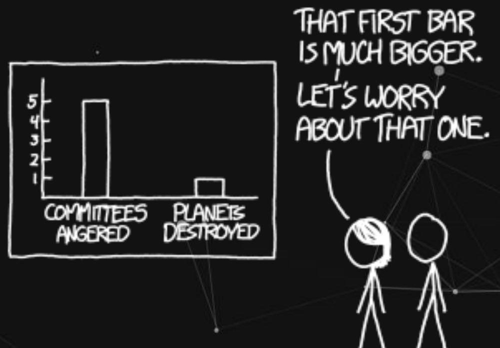
	Empowered	Unempowered
Ethical	😎	🤔
Unethical	😈	🤪

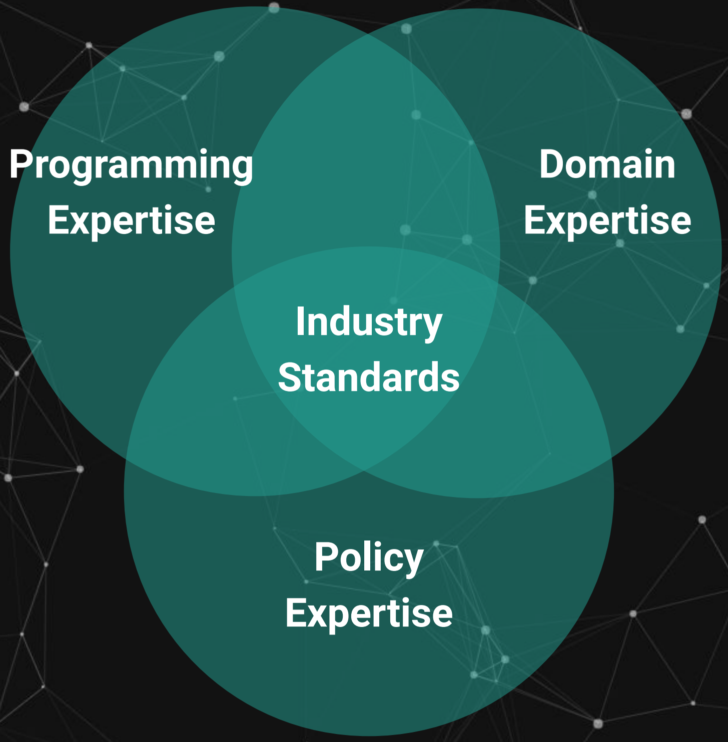
- Ethical ~~ Ought to do good
- Empowered ~~ Know how to

Going beyond the algorithms

@AxSaucedo

Large ethical challenges cannot fall on the shoulders of a single software developer





@AxSaucedo

End-to-end Approach

Industry standards & regulatory frameworks

2 Practical guidelines that set the bar for requirements around risk assessment and evaluation for machine learning systems

Principles & Guidelines

1 High level guidelines that provide a principled approach towards designing, building and operating machine learning.

Open Source Software

3 Practical implementations of the best practices on the infrastructure that provides the backbone to most applications.

Terminology

@AxSaucedo

Ethics

Moral principles that govern a person's behaviour or the conducting of an activity.

Principles

Fundamental truths or propositions that serve as the foundation for a system of belief or behaviour or for a chain of reasoning.

Why not just follow existing rules?

When dealing with new technologies/situations, there may just not be enough examples to base on, but practitioners will need to make decisions

Whose Ethics?

@AxSaucedo

Eastern? Western? ...?

The individual, continuity, good, the righteous, ...

Philosophical Foundations

!=

Current (Geo)political ecosystem

Understanding underlying philosophical foundations allows us to understand where we come from, to come to more powerful mutual agreements

Principles & Ethics Framework

@AxSaucedo

ethics.acm.org

Professional Ethics in Computing

acm

The ACM's Code of Ethics & Professional Conduct

1. Human Augmentation

I commit to assess the impact of incorrect predictions and, when reasonable, design systems with human-in-the-loop review processes

2. Bias Evaluation

I commit to continuously develop processes that allow me to understand, document and monitor bias in development and production.

3. Explainability by design

I commit to develop tools and processes to continuously improve transparency and explainability of machine learning models where reasonable

4. Reproducible systems

I commit to continuously improve my machine learning infrastructure to enable for a reasonable level of reproducibility

5. Displacement Strategy

I commit to identify and document relevant information so that business change processes can be developed to mitigate the impact towards workers being automated

6. Practical Accuracy

I commit to develop processes to ensure my accuracy and cost metric functions are aligned to the domain-specific applications

7. Trust beyond the user

commit to build and communicate processes that protect and handle data with stakeholders that may interact with the system directly and/or indirectly

8. Data risk awareness

I commit to develop and improve reasonable processes and infrastructure to ensure data and model security are being taken into consideration during the development of machine learning systems

The IEML's Principles for Responsible AI ethics.acm.org

Professional Ethics in Computing

acm

@AxSaucedo

Contribute to society and to human well-being...
Avoid harm
Be honest and trustworthy
Be fair and take action not to discriminate
Respect the work required to produce new ideas...
Respect privacy
Honor confidentiality

Strive to achieve high quality...
Maintain high standards...
Know and respect existing rules...
Accept and provide appropriate professional review
Perform work only in areas of competence
Foster public awareness and understanding...
Access computing and communication resources only when authorized
Design and implement systems that are robustly and usably secure

Principles = good for business

and software!

Industry/Code Standards

@AxSaucedo

Who sets code/industry standards?

You!

Who uses the industry standards?

Maybe **You!**

and maybe them too...

Standard: A repeatable, harmonised, agreed & documented way of doing something

Standardisation Bodies

You can get involved in the design and development and use of standards

ISO

IEEE

python SOFTWARE FOUNDATION

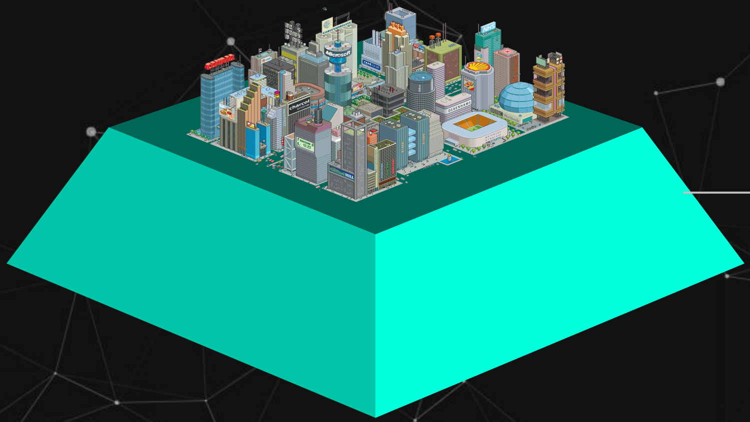
W3C

@AxSaucedo

Open Source as Foundation

@AxSaucedo

Open source is now ~~becoming~~ the backbone for critical infrastructure that runs our society



Open Source Software

3

Practical implementations of the best practices on the infrastructure that provides the backbone to most applications.

Open Source as Policy

@AxSaucedo

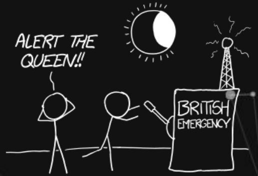
Principles are **useless** if the foundation is not in place to introduce and manage

Principles & Guidelines

1 High level guidelines that provide a principled approach towards designing, building and operating machine learning.

Open Source Software

3 Practical implementations of the best practices on the infrastructure that provides the backbone to most applications.



Open Source as Lead

@AxSaucedo

Open source leaders are developing the core cogs that regulation depends on

Principles & Guidelines

1

High level guidelines that provide a principled approach towards designing, building and operating machine learning.

Open Source Software

3

Practical implementations of the best practices on the infrastructure that provides the backbone to most applications.

Open Source Foundations

You can get involved on the design and development and use of standards

@AxSaucedo

python SOFTWARE FOUNDATION

THE LINUX FOUNDATION

software freedom conservancy

THE APACHE™ SOFTWARE FOUNDATION

Sidenote: Regulation

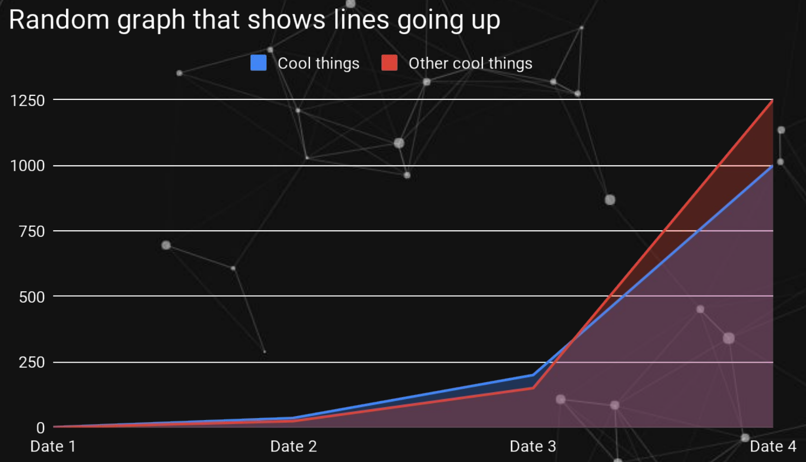
@AxSaucedo

We all can agree: Bad regulation is BAD.

However good regulation can be a catalyst for innovation through enforcement of best practices and mitigation of bad actors.

Software's Massive Traction

@AxSaucedo



Growth

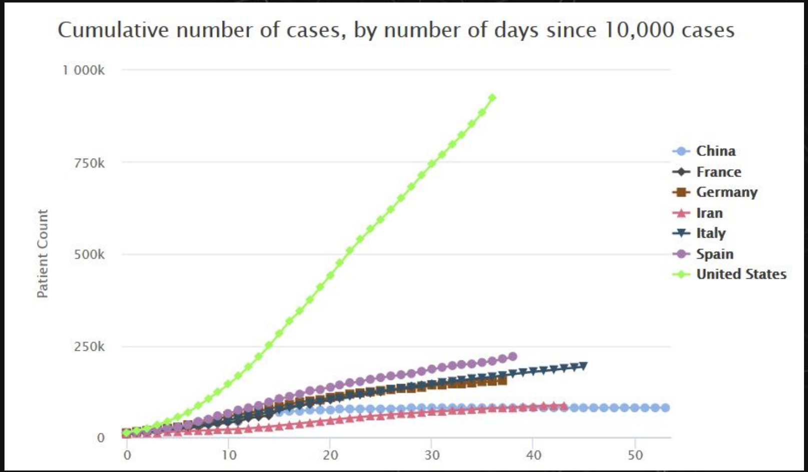
- Internet Services
- Machine Learning Automation
- Cloud Native infrastructure
- Gaming and design tools
- Etc, etc, etc, etc

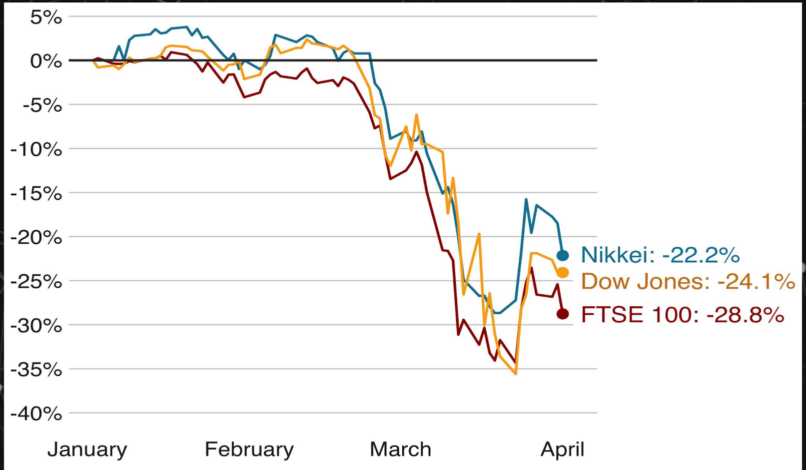
@AxSaucedo

Not all can be solved w code

When you run around with a **hammer** everything may look like a **nail**

E.g The Challenge of our Generation

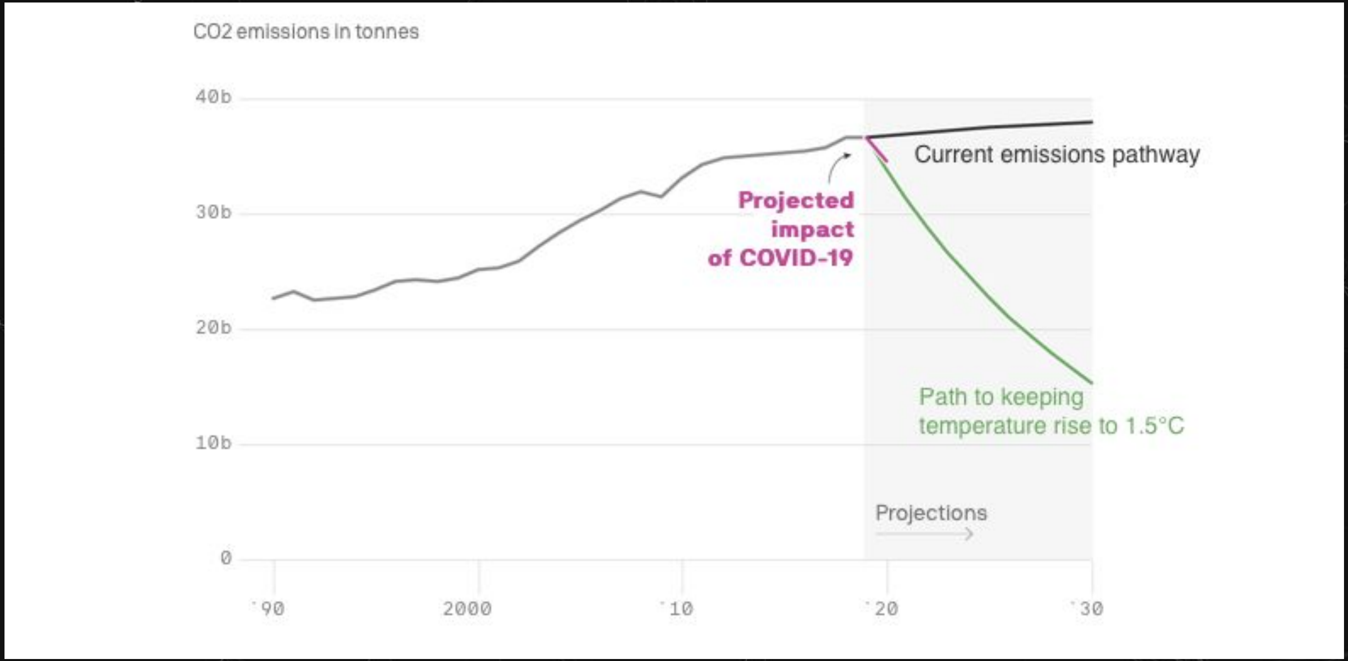




Societal Impact

Economic Impact

And potentially not the last



https://medium.com/@amynoelle/flatten-the-climate-change-curve-2ed756eaa082

@AxSaucedo

Ensuring the right solution

@AxSaucedo

Before tackling a problem we should be able to identify how much of it is actually a software problem before actually writing code

And whether the solution is even solving a problem

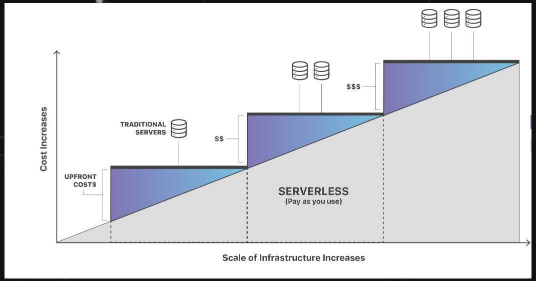
@AxSaucedo

Practical Deep Dive

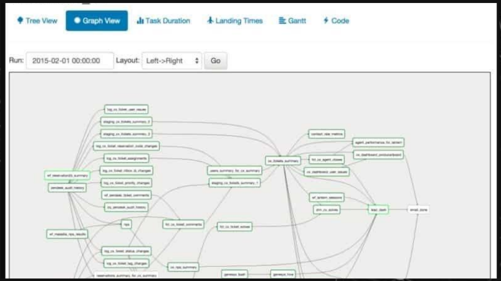
Production machine learning systems

Prod ML Systems are HARD

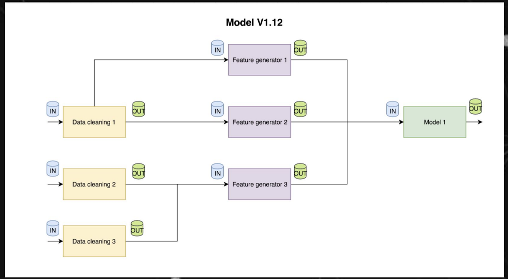
@AxSaucedo



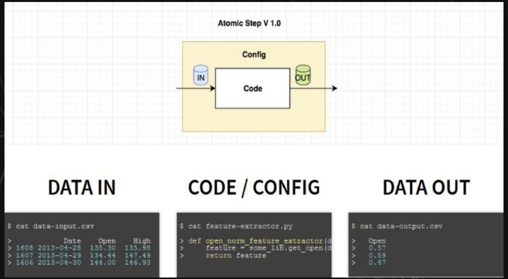
Specialised Hardware (GPU, etc)



Complex Dependency Graphs



Compliance



Reproducibility of components

Last year's talk on the challenges & landscape in ML: https://www.youtube.com/watch?v=Ynb6X0KZKxY

Principles for responsible AI

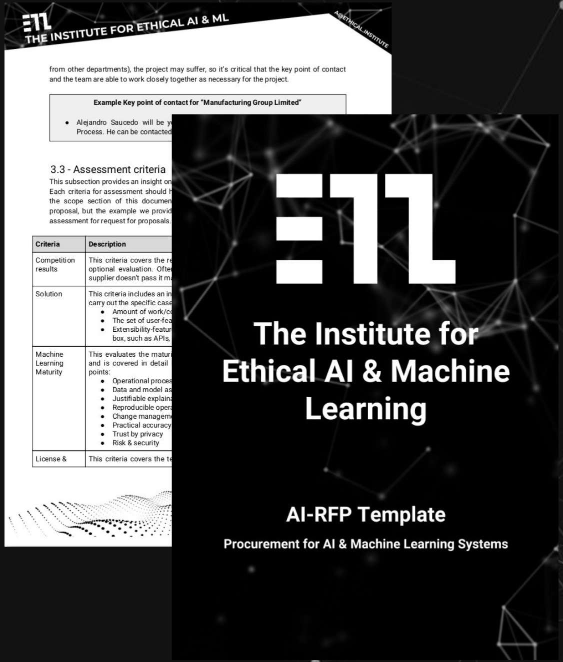
@AxSaucedo

1. Human augmentation / review
2. Bias evaluation capabilities
3. Explainability by justification
4. Reproducible ops infrastructure
5. Displacement strategy
6. Practical statistical metrics
7. Trust by privacy
8. Security risks

http://ethical.institute/principles.html

Procurement Framework

@AxSaucedo



THE INSTITUTE FOR ETHICAL AI & ML

A@ETHICAL.INSTITUTE

from other departments), the project may suffer, so it's critical that the key point of contact and the team are able to work closely together as necessary for the project.

Example Key point of contact for "Manufacturing Group Limited"

- Alejandro Saucedo will be y… Process. He can be contacted…

3.3 - Assessment criteria

This subsection provides an insight on… Each criteria for assessment should h… the scope section of this documen… proposal, but the example we provid… assessment for request for proposals.…

Criteria	Description
Competition results	This criteria covers the re… optional evaluation. Ofte… supplier doesn't pass it m…
Solution	This criteria includes an in… carry out the specific case… • Amount of work/c… • The set of user-fea… • Extensibility-featur… box, such as APIs,…
Machine Learning Maturity	This evaluates the maturi… and is covered in detail… points: • Operational proces… • Data and model as… • Justifiable explaina… • Reproducible opera… • Change managem… • Practical accuracy… • Trust by privacy • Risk & security
License &	This criteria covers the te…

The Institute for Ethical AI & Machine Learning

AI-RFP Template

Procurement for AI & Machine Learning Systems

A set of templates for industry practitioners:

- Request for proposal
- ML maturity model
- Tender competition template

http://ethical.institute/rfx.html

ML Maturity Model

@AxSaucedo

Practical benchmarks
Explainability by justification
Infrastructure for reproducible operations
Data and model assessment processe
Privacy enforcing infrastructure
Operational process design
Change management capabilities
Security risk processes

From principles to a checklist

- Each has a set of questions for supplier compliance
- Top-bottom approach providing red flags

http://ethical.institute/rfx.html

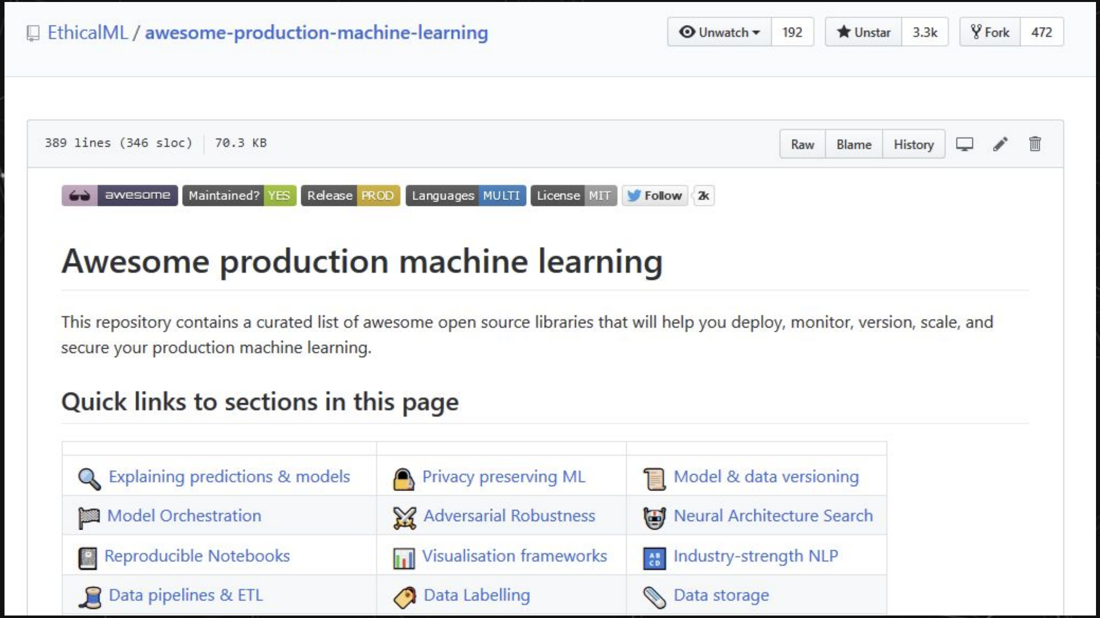
Alignment on first principles

@AxSaucedo

#1	Supplier doesn’t have infrastructure and/or processes to version different machine learning models where reasonable
#2	Supplier does not have a protocol to evaluate whether new ML model requires domain expert for evaluation of low confidence results
#3	Supplier system doesn’t have capabilities to perform development across production and QA/BETA environments
#4	Supplier does not have a process and/or infrastructure to revert models in production without unreasonable level of disruption
#5	Supplier doesn’t have processes and/or infrastructure that ensures only users with explicitly granted permissions have access to PII data
#6	Supplier doesn’t have process to assess human review process requirements based on the impact of incorrect predictions
#7	No process and/or infrastructure to ensure machine learning data encrypted on transport/rest
#8	Supplier doesn’t have a process and/or infrastructure to introduce specialised model evaluation metrics where required

http://ethical.institute/rfx.html

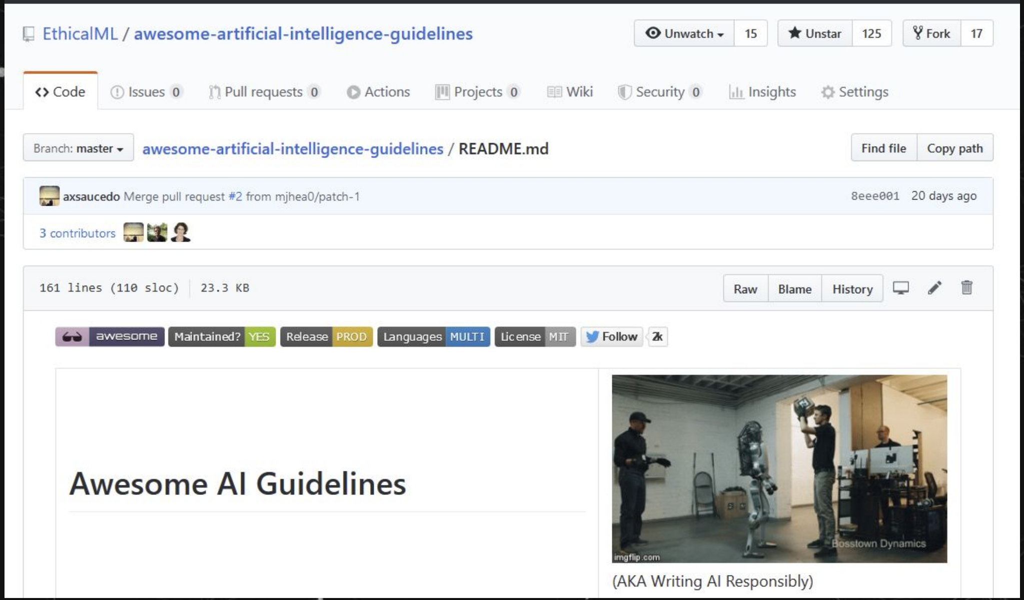
Broader list of Prod OSS libraries



http://bit.ly/awesome-mlops

Broader list of guidelines

@AxSaucedo



https://github.com/EthicalML/awesome-artificial-intelligence-guidelines

Industry Framework Case Study

@AxSaucedo

CORE	#2 Bias evaluation #3 Explainability
SECONDARY	#8 Security #1 Human-in-the-loop #6 Practical metrics

Loan approval process

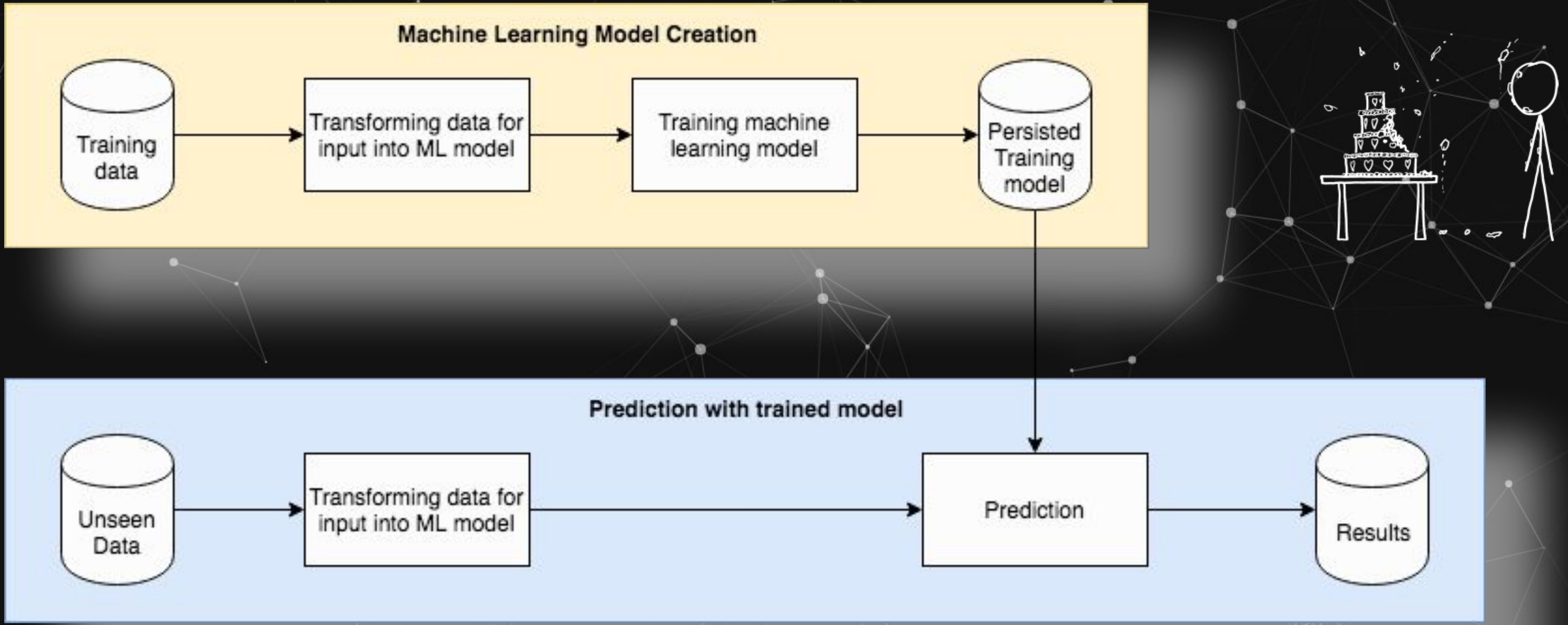
Domain expert evaluates application
Loan is approved or rejected
Manual process

Business wants to automate this
process with machine learning

@AxSaucedo

Traditional data science proces

@AxSaucedo



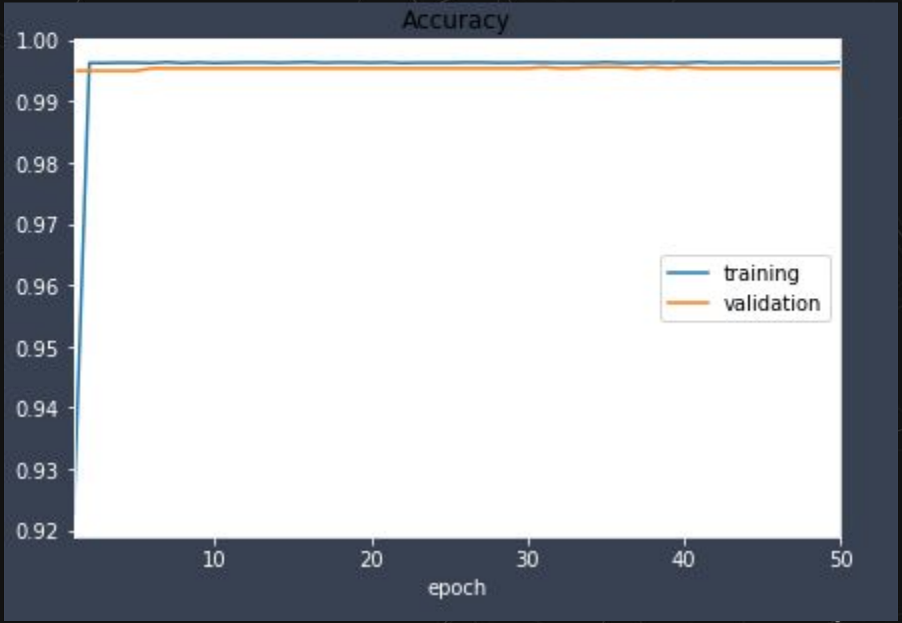
@AxSaucedo

We obtain some data

age	workclass	education	education-num	marital-status	occupation	relationship	ethnicity	gender	capital-gain	capital-loss	hours-per-week	native-country	loan
39	State-gov	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	False
50	Self-emp-not-inc	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	False
38	Private	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	False
53	Private	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	False
28	Private	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	False

We get 8000 rows with target

We train our model

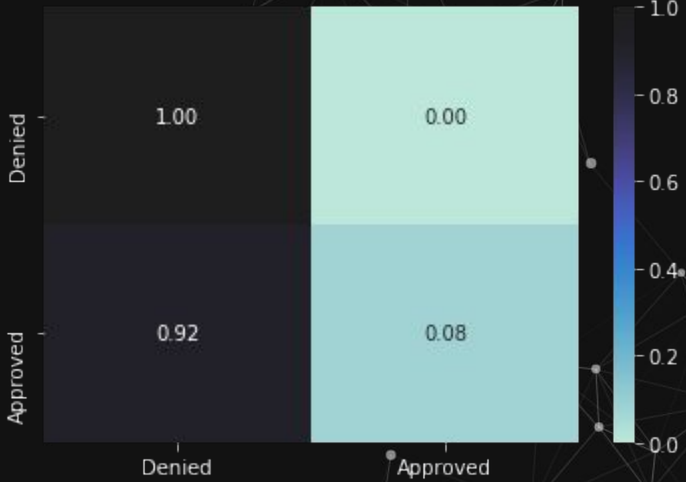


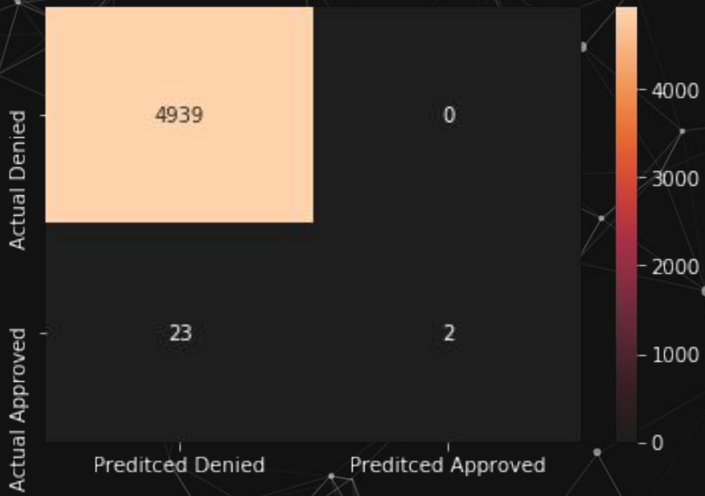
99% Accuracy

Time for production?

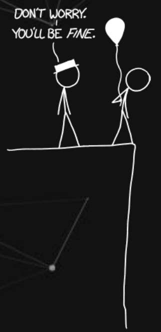
It’s a disaster

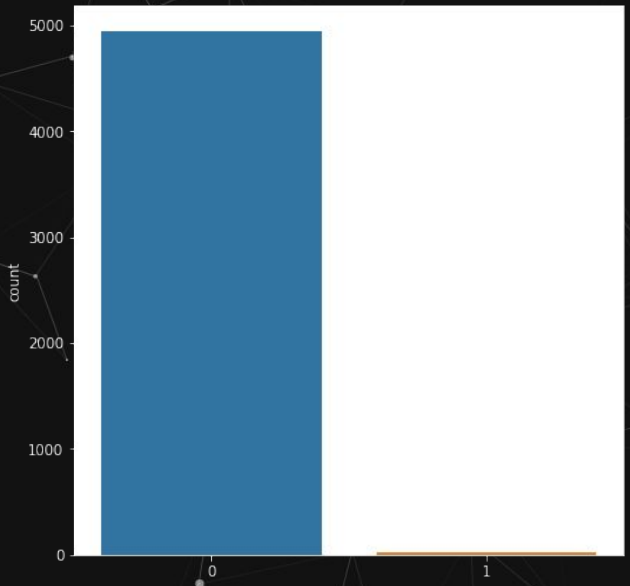
@AxSaucedo



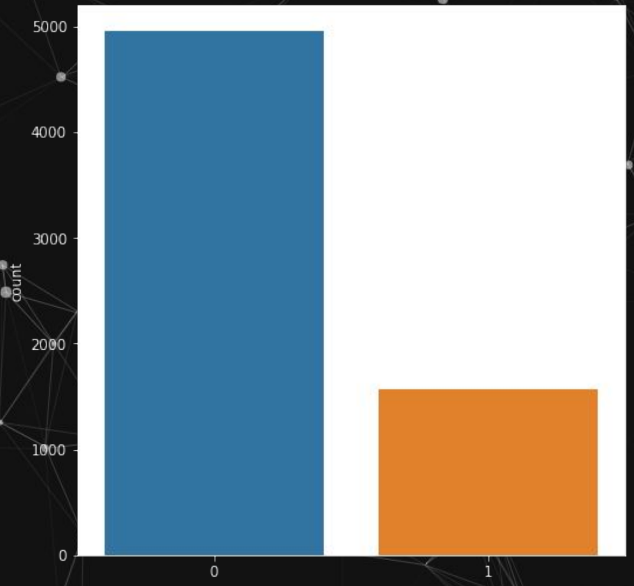


When we look at our data...



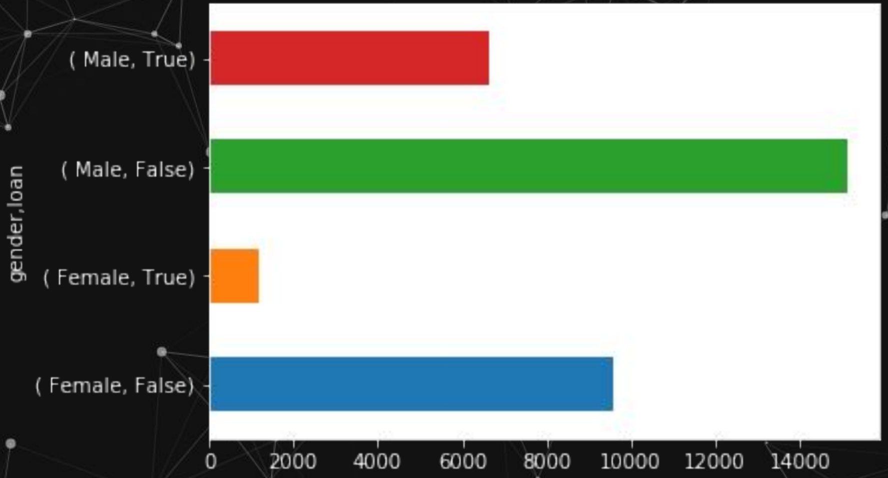


Training data



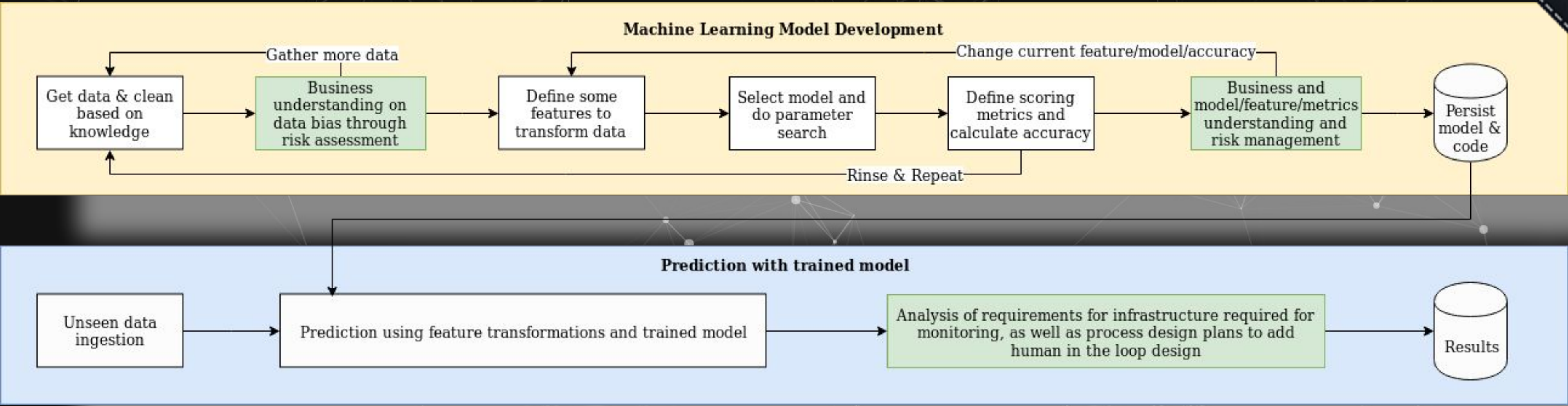
Production data

Let’s analyse dataset further



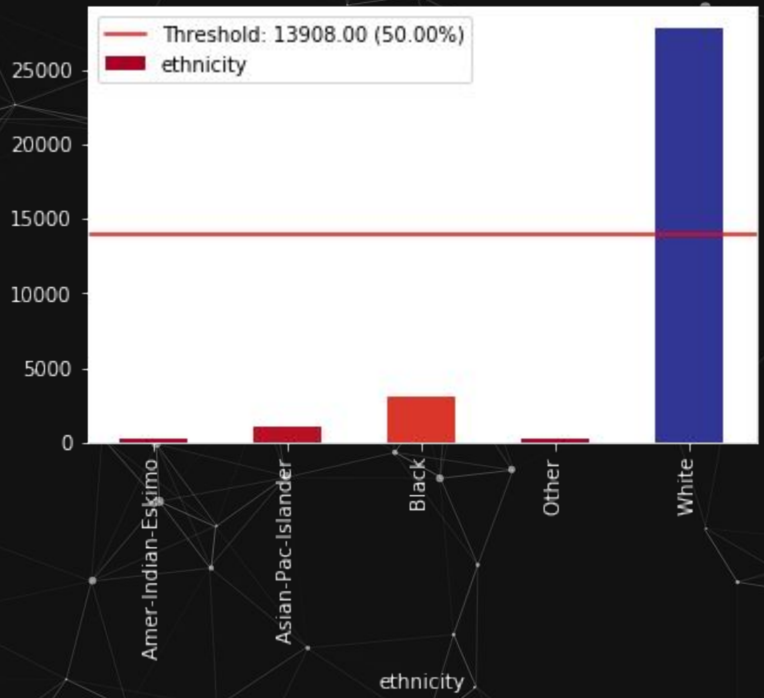
@AxSaucedo

Bias Evaluation Process

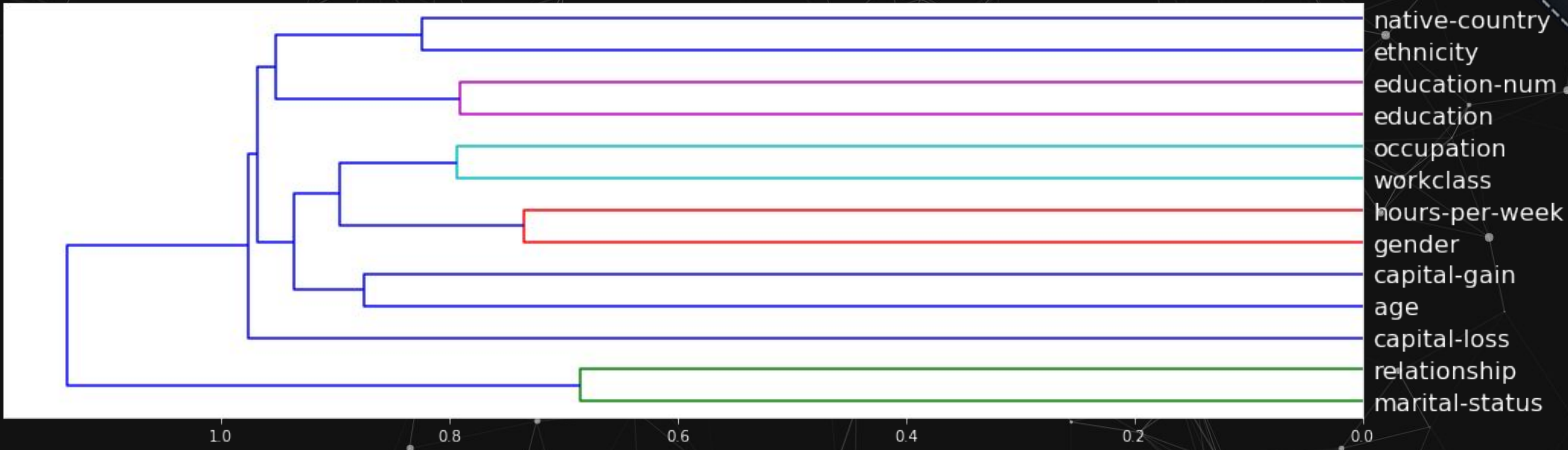


@AxSaucedo

We can upsample/downsample

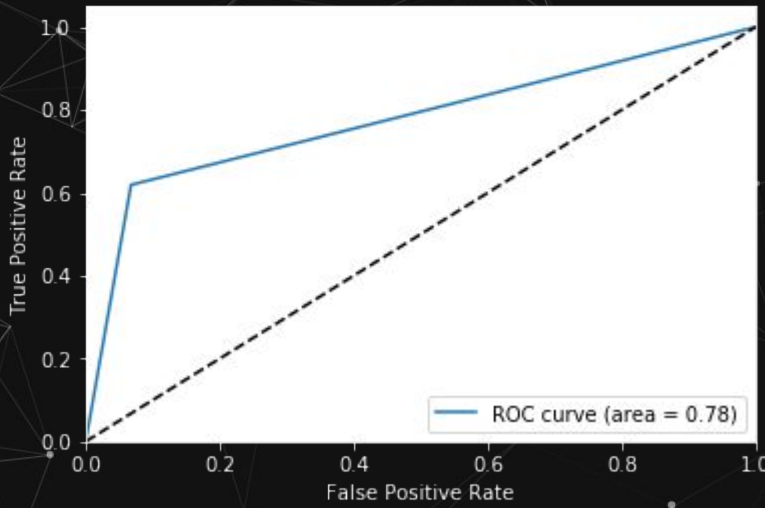


Taking into account correlation



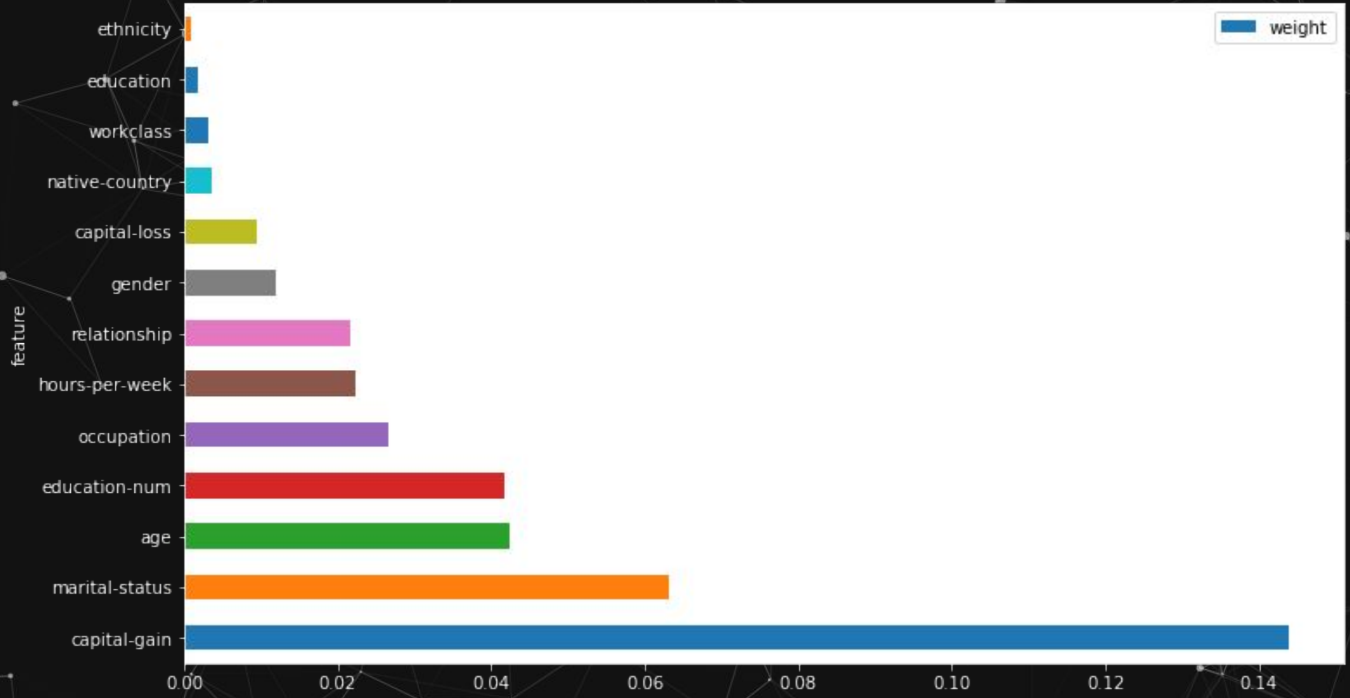
Much better...

0.93
0.07
0.38
0.62
Denied
Approved
Denied
Approved
0.90
0.75
0.60
0.45
0.30
0.15



True Positive Rate
False Positive Rate
ROC curve (area = 0.78)
0.0
0.2
0.4
0.6
0.8
1.0

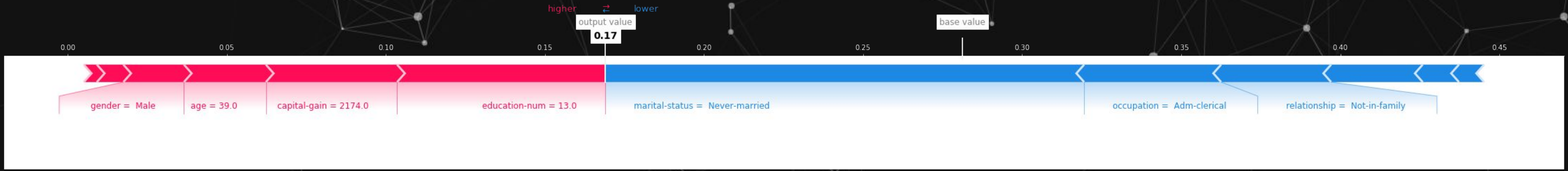
Let's explain predictions



@AxSaucedo

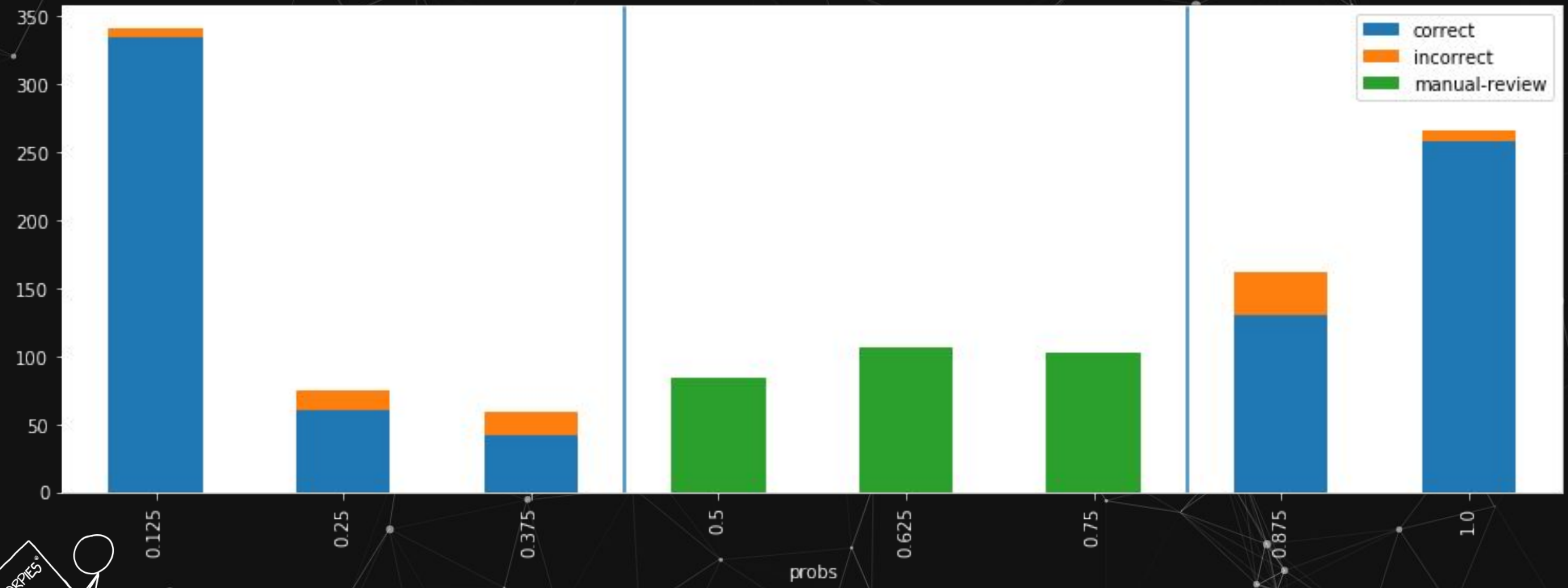
Let’s explain predictions

@AxSaucedo



We can add manual review

@AxSaucedo



@AxSaucedo

Recap

The impact of software development
Responsibility as individual and organisations
Ethics and Principles
Industry & Code Standards
Finding the right solution for the right problem
Practical deep dive on AI

Meditations on First Deployment

A Practical Guide to Responsible Development

EuroPython 2020

Alejandro Saucedo

@AxSaucedo

Massive Shoutout to

what-if.XKCD.com

For their always-amazing artwork & content!
Check it out and support them!

@AxSaucedo