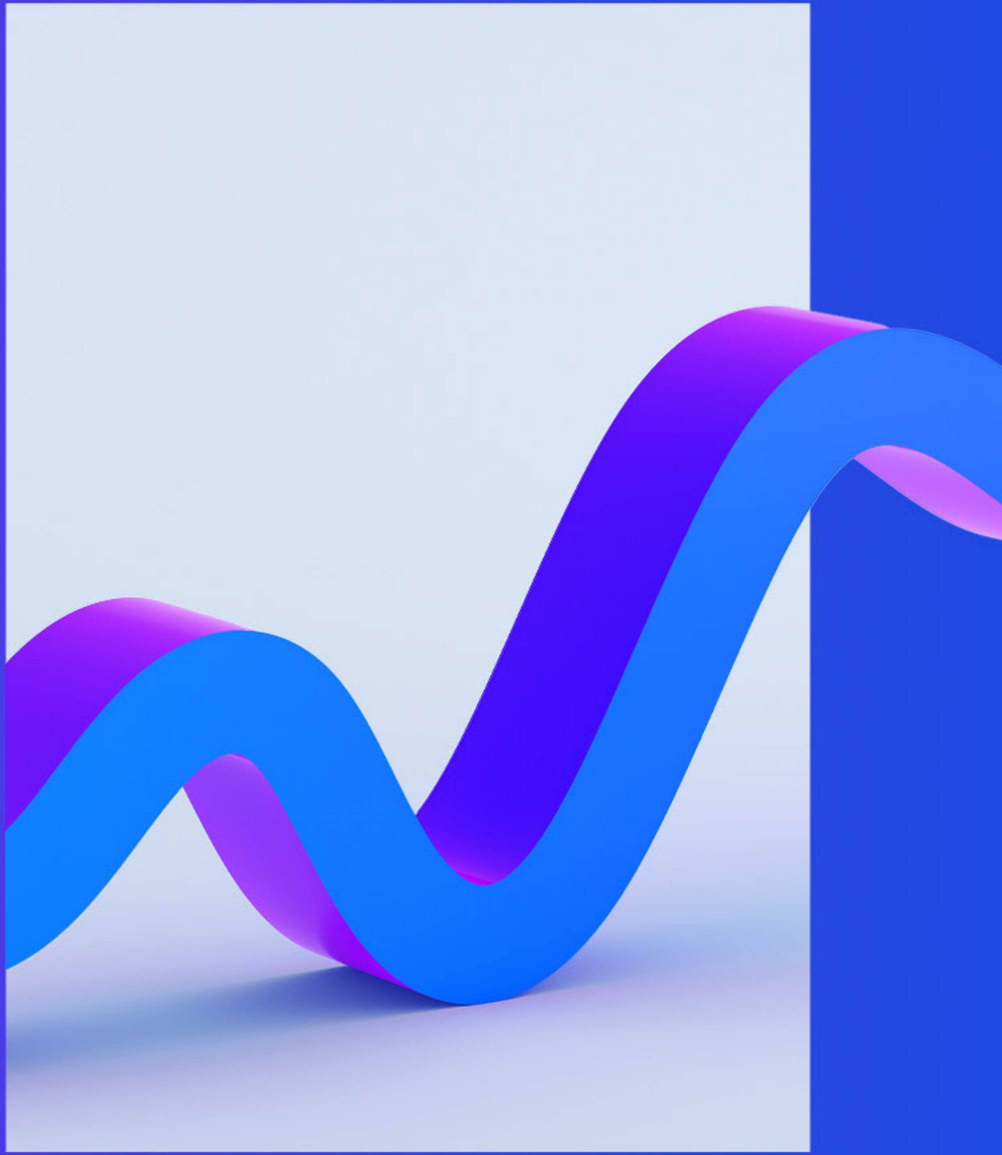
Modern Risk Management for AI Models - Making AI Responsible



Re-imagining the Model Risk Management function for Artificial Intelligence / Machine Learning models

November 2022

Your KPMG-team:
Today's hosts and speakers are ready to go

Hosts and Core Team

Matthias Peter
Partner
KPMG in Germany

Rajosik Banerjee
Partner
KPMG in India

Janek Gallitschke
Senior Manager
KPMG in Germany

Dr. Christoph Anders
Manager
KPMG in Germany

Your Speakers

Kinshuk Pal
Associate Partner
KPMG in India

Dr. Philipp Prasse
Assistant Manager
KPMG in Germany

Dr. Alina Braun
Assistant Manager
KPMG in Germany

Welcome & Introduction
Main Speaker
Theory / Background
Hands on / Prototype

With important contributions by the following KPMG colleagues:

Daniel Wehl,
Fabian Slezak,
Leon Schmitz

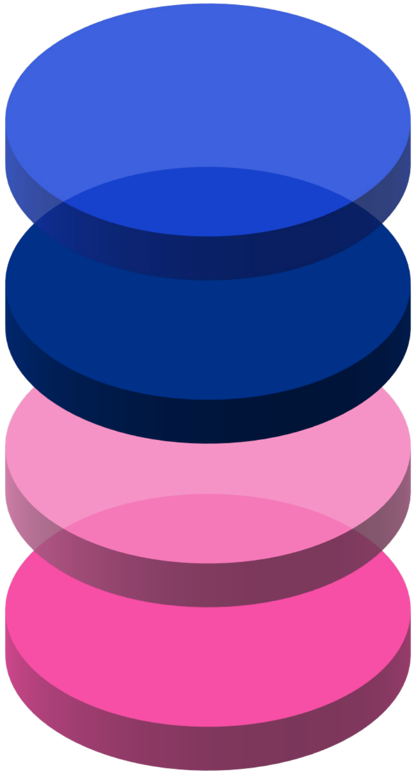
There is no doubt about the potential of Machine Learning (ML) – It just needs to be leveraged in banks

“New technologies such as AI [Artificial Intelligence] and machine learning offer tremendous opportunities for both banks and supervisors.

However, to use these technologies **safely and soundly**, we need an **adequate regulatory framework, proper supervisory oversight and** an **understanding by all users** – banks and supervisors alike – of not just the potential but also the **limitations and risks** of these technologies.”

Elizabeth McCaul
Member of the Supervisory Board of the ECB, July 2022
Speech at the conference on “The use of artificial intelligence to fight financial crime”, organized by Intesa Sanpaolo

The quote is a great starting point for today's topic: AI / ML in Model Risk Management with an additional focus on fairness



01 Risk Management of AI / ML models
- Why it is Important?

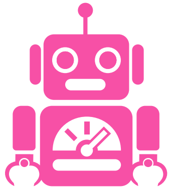
02 Modern Model Risk Management
- Re-imagining the MRM framework

03 Deep-dive Fair AI / ML models
- New challenges for model validation

04 Deep-dive Fair AI / ML models
- Mitigation methods, fairness pipeline

Based on KPMG Whitepaper

Use case
- Credit scoring



The management of AI / ML risks is gaining importance due to rising public and regulatory attention

High potential - high challenges

- AI / ML usage is widespread and becoming the norm in many industries
- In banking an increasing use of AI / ML can be observed
- Examples are: Self-driving-finance, fraud detection, data analytics, deep hedging
- Use of AI / ML comes with both advantages and specific risks
- The specific risks must be taken into account when using ML

Increasing relevance for bank management

- Increased use of AI/ML in various areas: Customer acquisition / retention, pricing, data management, compliance & fraud, risk management
- Applications can be found in less regulated areas due to large regulatory uncertainties

Increasing public interest

- Machine learning is increasingly used in direct relation to the customer
- AI decisions might increase the risk of negative and harmful impact on private persons

Increasing Regulatory requirements

- Many regulatory publications on European and national level – However, no actual regulatory requirements as of yet
- Specific and additional regulation w. r. t. governance, model risk management, and running of AI / ML can be expected

AI Specifics

- Traditional MRM processes are often not capable to address specific risks of AI / ML models and regulatory requirements
- In particular model choice, parametrization / feature engineering, explainability, and fairness are challenging

Machine Learning is applied in banks all the way from front to back office

Select use cases

Customer Acquisition / Retention

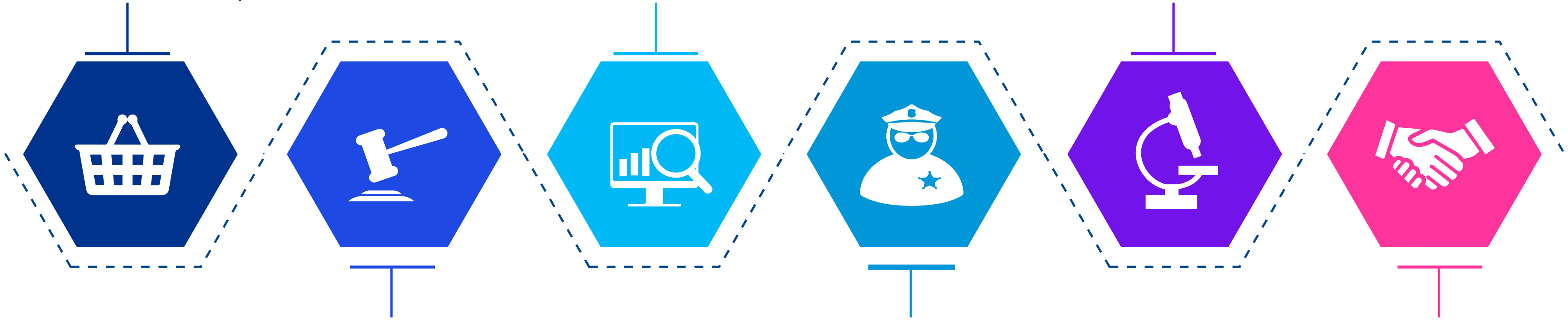
- Individualization of customer offerings incl. cross-selling
- „Self-driving finance“ & improvement of customer experience
- Identification of new market potential

Data Management

- Improvement and automatization as well as testing of data quality
- Optimization of the internal / regulatory reporting in terms of content and processes

Risk Management

- Better calculation methodology
- Improvement of data quality
- Intraday capability
- Process optimization, validation



New Business

- Improved customer information incl. KYC
- Better business decisions (credit scoring, pricing)

Compliance & Fraud

- Identification of money laundering (AML)
- Detection of account / credit card fraud
- Cybersecurity support
- Monitoring of retail activities

Market Infrastructure

- Optimization of “post-trade” processes
- Trade execution improvement (in unstable markets)

Growing public interest due to prominent mistakes of AI algorithms call for an increasing need for transparency

The use of AI involves risks

Apple Card Investigated After Gender Discrimination Complaints

Amazon scraps secret AI recruiting tool that showed bias against women

SELF-DRIVING CARS —
How terrible software design decisions led to Uber's deadly 2018 crash

Increasing need for transparency, regulation, and risk management

KPMG study

94%

The large majority (94%) of participants expect AI governance challenges to be carefully managed.

57%

Most participants would be **more willing to use AI** systems if assurance mechanisms were in place, such as independent **AI ethics reviews**, **AI ethics certifications**, and AI codes of conduct.

81%

The large majority of participants (81%) expect AI to be regulated.

Source: KPMG Study - Trust in Artificial Intelligence: A five country study (2021) LINK

Due to the inherent risks of AI / ML there is an increased regulatory focus on the use of AI / ML methods

Regulatory requirements

- Various regulatory publications regarding the use of AI / ML in the EU – Currently no applicable laws
- Latest publications by BIS and Bank of England on validation expanding SR 11-7
- Long-term impact on model risk management frameworks to be expected

EIOPA
AI Governance principles

BaFin
Principles for the use of algorithms in decision-making

BoE
Model Risk Mgmt. principles for banks (consultation)

2021

2022

European Com.
Proposal for a Regulation for AI
(EU AI Act)

BaFin/ Bundesbank
Machine Learning in risk models (consultation)

EBA
Discussion paper on machine learning for IRB models

Key Takeaways

Industry agnostic EU regulation is also relevant for Banks

- Broad definition of ML includes all statistical methods, i.e. logistic regression
- Credit scoring is explicitly mentioned as an example for high risk uses that have to be treated especially rigorous

Currently no actual regulatory requirements

- Publications are on the level of drafts, discussion / principle papers – Not actual laws
- Requirements / demands for pillar I continue to be unclear

Focus: Explainability, fairness and accountability

- All publications focus on ML specific topics that are partially new for banks
- Key topics are explainability, fairness, and accountability

Focal points for AI governance can be derived from regulatory publications and the specifics of AI / ML

	Requirement	Newness	Complexity
Adequacy	**Similar requirement as traditional models, but:** AI / ML models require new approaches to validation, stronger focus on data and stronger ongoing monitoring	BCBS (e.g. risk data aggregation & risk reporting), Basel Core Principles, CRR II/III, TRIM	Due to high complexity of the model and specific model cycle
Transparency / Explainability	**Explainability of the method is one of the most critical issues in AI / ML:** Application of new methods such as XAI - Explainable AI necessary. Approaches require know-how building and new technical solutions	Only a few requirements in existing regulations (CRRII/III)	Machine Learning Algorithms i.e. black boxes
Fairness, ethics	**High social relevance - Currently not sufficiently taken into account:** Intensive research and further developments in the topic to be observed. Requires a new approach to data, methods, and results	No consideration in previous regulatory framework for banks	Front-to-back to be considered, no empirical values, imprecise specifications.
Accountability	**Additional requirements in addition to those on traditional models:** **Human-in-the-Loop:** Human influence in decision-making **Human-on-the-Loop:** Human influence in design and review	BCBS (Corporate governance principles for banks), Basel Core Principles, CRR II/III	Processual effort with resource utilization incl. documentation
Data privacy, third party	• **Data privacy:** Ensuring privacy in all steps of the processing, if necessary, enquiry about the use of the data for training • **Third party:** Same requirements as for in-house applications	Extensive detail and regulation through DSGVO	New customer communication and data protection concepts necessary

Traditional MRM processes are often not capable to address specific risks of AI / ML models and regulatory requirements

AI Models need a comprehensive MRM Framework

- The basis for regulatory compliance is a working model risk framework
- However, existing model risk frameworks are usually not suited for AI / ML model specific challenges
- There are several challenges that need to be taken into account
- Those challenges require specific adaptation of exiting model risk frameworks

Select AI / ML specific challenges for Model Risk Governance

Models are more opaque and so require coordination between model developers, users and validators

Skillset requirements are different from tradit-ional MRM processes

Some **use cases are not models as per SR 11-7** definitions

Avoiding bias and ensuring fairness and ethics in the model

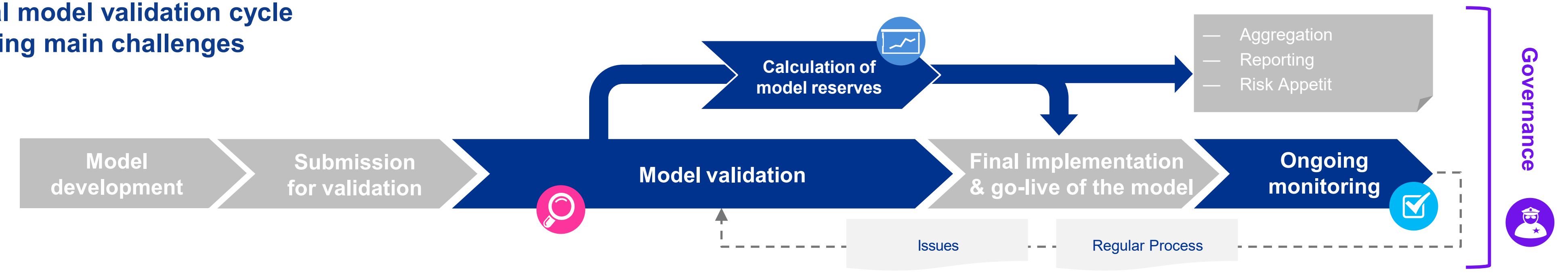
ML models undergo recalibrations more often than traditional MRM processes

Assumptions of static performance between two MRM reviews are most often not true

In particular, along the entire model validation cycle, the challenges of Machine Learning models have to be considered.

The validation approach and processes need to be adapted to and extended for Machine Learning models

Typical model validation cycle including main challenges



Challenges for the different validation components

Model Validation (Deep-Dive next Slide)	Ongoing Monitoring	Model reserves	Governance
Underlying data	Re-Training	Benchmark models	Responsibilities
Model choice	Model changes	Risk approximation	Documentation
Parametrization / Feature Engineering	Explainability	Monitoring	Certificates
Fairness	Fairness		Open-Source / third party
Explainability			Fairness

The challenges faced when validating ML algorithms need to be met with the use of new methods

Model validation	Challenges
Underlying data	• Larger data sets, different data structure and content - validation regarding bias and fairness necessary • Ensuring representativeness of training and test sets for productive data
Model choice	• Review the appropriateness of the model in terms of model performance, explanatory power, fairness, and data basis • Validation feature selection from the raw data incl. "business backgrounds"
Parametrization / Feature Engineering	• Higher importance and larger number of hyperparameters in ML algorithms – "Nature" of parameter difference compared to classical models
Explainability	• Explainability of machine learning models not given or challenging for certain approaches • Use of Explainable AI required
Fairness (Deep-Dive next Slide)	• Fairness is partly a completely new topic for validation without defined responsibilities and know-how • Application of new methods required

Validation methods

Even for simple Machine Learning methods specific statistical methods or specific aspects have to be considered during validation.

Examples

Machine Learning Methods

- **Linear Discriminant Analysis:** Wilks lambda, function of group centroids, canonical structure matrix,
- **Decision Tree:** Splitting criteria, stopping criteria, root node, decision node
- **Deep Learning:** Hyperparameter (e.g. #Layer), backpropagation, loss function, activation function

Explainable AI

LIME, SHAP, Anchor, PDE, ICE, Counterfactual

Fairness is one of the biggest challenges in the application of ML algorithms besides XAI and requires a high level of attention and the application of new approaches.

Definition of fairness and its translation into mathematical formulae are among the biggest challenges for ML for banks

What is Fairness?

Definition of fairness varies depending on one's perspectives and circumstances and therefore no single definition is true for all cases[1], e.g.:

(Anti-discrimination) Law

Philosophy

Social sciences / Public opinion

The definition of fairness in these areas is usually very abstract. For application in ML, a translation into mathematical terms is necessary.

Quantitative fields

In **Quantitative fields f**airness is a mathe-matical problem where some sort of criteria need to be fulfilled (equal representation or error figures)

Challenges

- There is no right answer when it comes to defining fairness
- Different stakeholders have different understandings of fairness (client, management, regulator, different cultures)
- Translation into mathematical formulae of a fairness definition is not always clear and can lead to the loss of nuances
- Selecting a fairness definition means making trade-offs - and these trade-offs need to be documented and understood
- Unfairness can arise not only from the model but also from the use or user of the model itself
- Improper use of ML results in the reproduction of or even an increase in bias

Approaches to ensuring fairness

Quantitative Measures (examples)

- Definition & use of different fairness metrics
- Pre-processing methods - adjustment of the training database
- In-process methods – adjustment of the model itself that is used for learning
- Post-Processing – adjustment of the results of the machine learning algorithm

Qualitative Measures (examples)

- Documentation of the model and the decisions made during development
- Checklists with different generally accepted fairness criteria
- Instructions for operation, user training, sensibilization of developers and users

[1] **This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology**. Mulligan, Kroll, Kohli Wong. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 119 (November 2019)

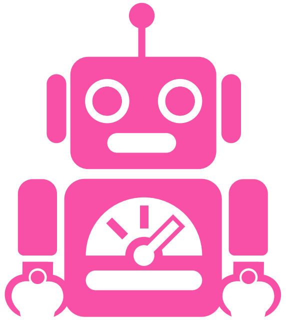
The challenges and methods for fair Machine Learning can best be demonstrated by a case study

Idea — **Illustration of fundamental approaches** to mitigating bias in ML for credit scoring.

Build — Use of a **publicly available dataset** and **implementation** of the Fair AI methods in **python.**

Result — **Step by step** presentation of results using simple but understandable methods.

Example



Credit Scoring

The purpose of this case study is not to give a comprehensive presentation of all methods and measures, but to provide a basic understanding of the idea behind the measures.

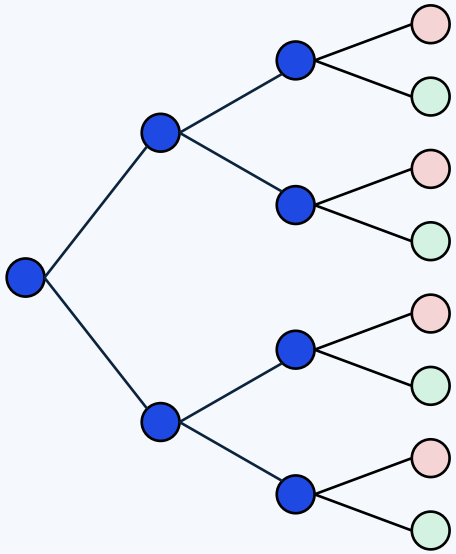
Making the fairness problem tangible by analyzing gender bias in credit scoring data

The data set: Test subjects John & Anna

	John	Anna
Age	67	32
Sex	Male	Female
Housing	Own	Own
Savings	Moderate	Little
Credit (in $)	1,169	13,832
Duration	6	48
Purpose	Radio/TV	Car
Risk	Good	Bad

Extract of the underlying data set. Relevant sensitive feature Sex (male / female) and relevant categorization Risk (good / bad).

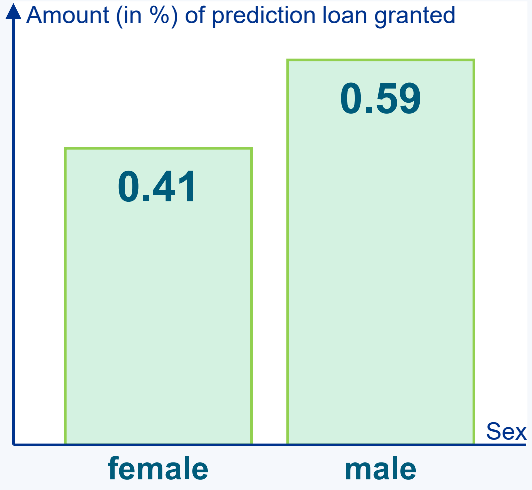
The classifier



Visualization of a random forest trained on the data set. Predicts if a person should be granted a loan (good Risk) or not (bad Risk).

The prediction

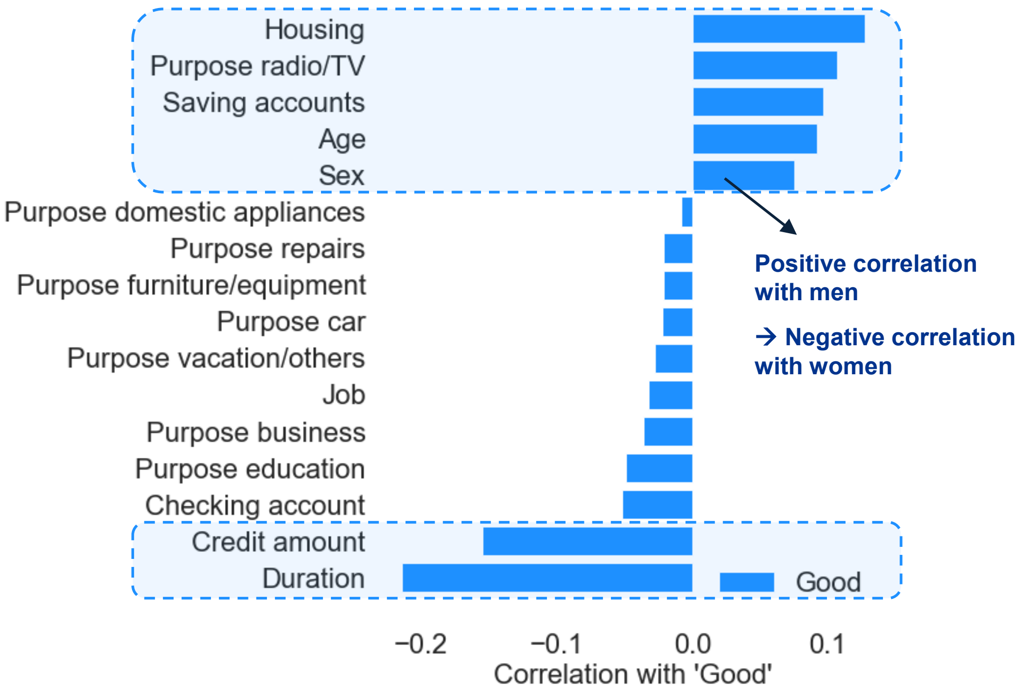
Proportion of women and men who are predicted to be granted a loan



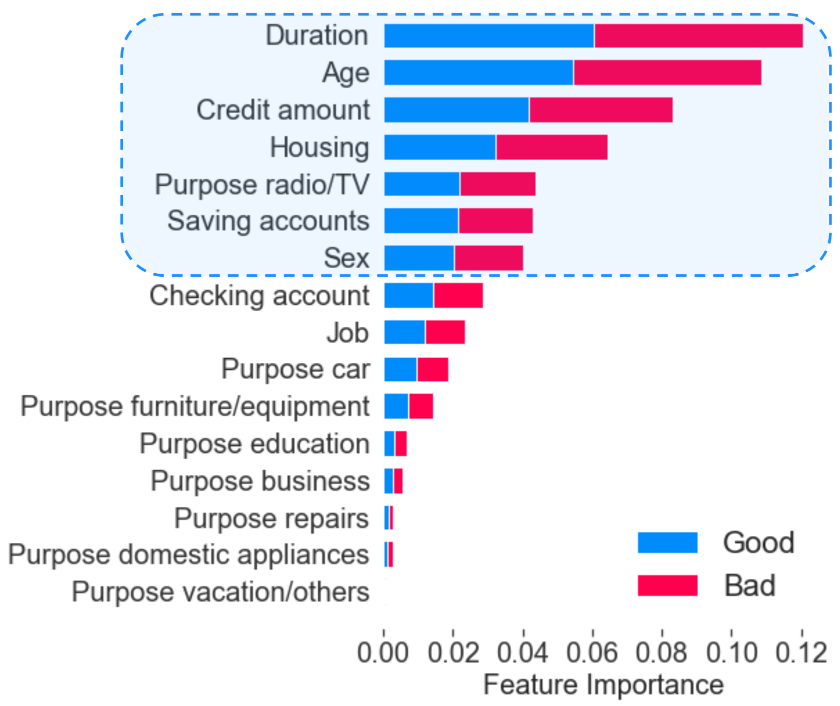
Visualization of the classification result. Shows the percentage of women and men among all who are predicted as good Risk.

Understand your data to gain insights into the learning algorithm and fairness

Data analysis: Correlation of input data



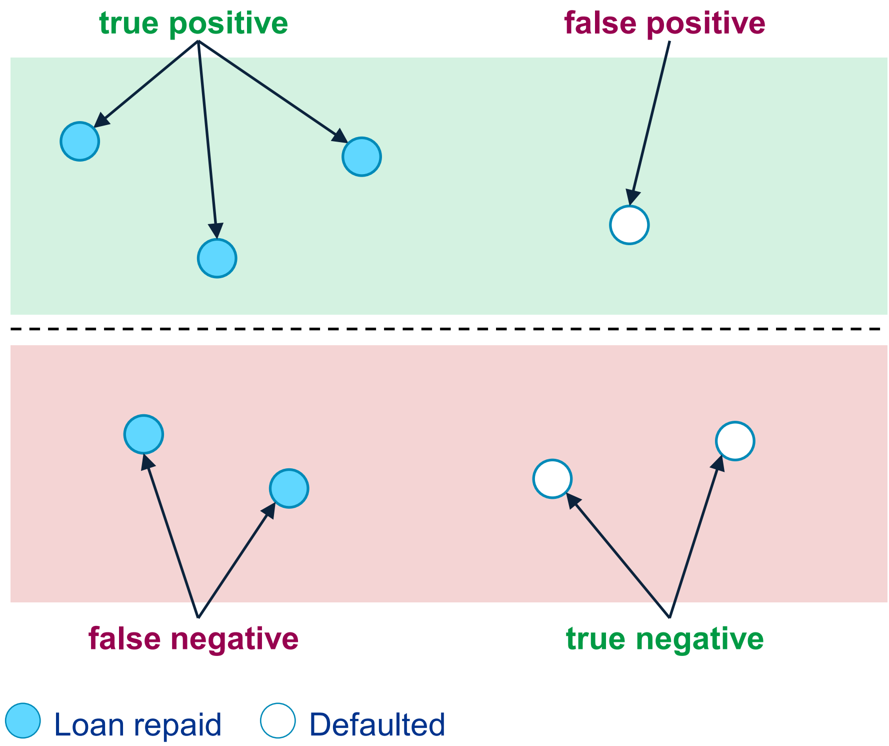
Model analysis: Feature importance



Should the **Gender** have a similar impact in the risk categorization as the **saving accounts**?

Approach to measuring fairness - The confusion matrix indicates the success and failure rates of the classification

Examples for prediction results



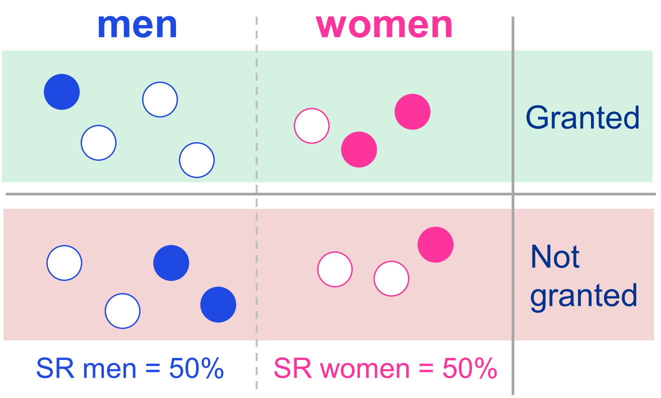
The confusion matrix

	Loan actually repaid	defaulted
Prediction Loan Granted	**true positive (TP):** **granted a loan to** a person that actually **would have repaid** the loan **# = 3**	**false positive (FP):** **granted a loan** to a person **that would** actually **not have repaid** the loan **# = 1**
Prediction Loan Not Granted	**false negative (FN):** **not granted a loan** to a person that actually **would have repaid** the loan **# = 2**	**true negative (TN):** **not granted a loan** to a person that **would** actually **not have repaid** the loan **# = 2**

There are different metrics to measure bias. Selection-, true positive und false positive rate are widely used examples

Equal selection rate (SR)

The chance of being selected by the model is equal for both groups (loan granted / not granted)

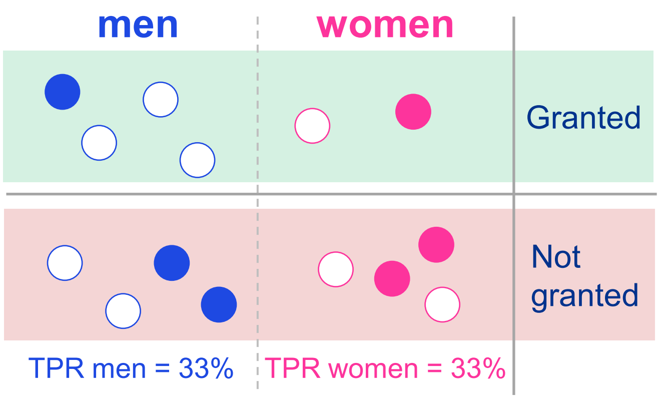


Formula

$$SR = \frac{TP + FP}{TP + TN + FP + FN} = \frac{2+1}{2+1+1+2} = 50\%$$

Equal true positive rate (TPR)

Given the condition of being qualified, the chance of getting a loan is equal for both groups (loan granted / not granted)

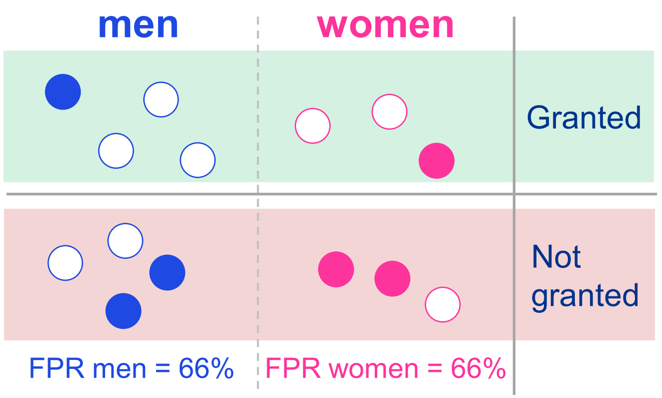


Formula

$$TPR = \frac{TP}{TP + FN} = \frac{1}{1+2} = 33\%$$

Equal false positive rate (FPR)

Given the condition of being unqualified, the chance of getting a loan is equal for both groups (loan granted / not granted)



Formula

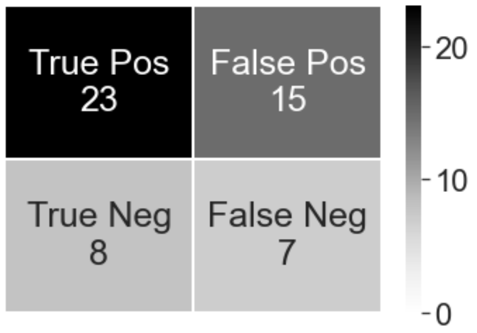
$$FPR = \frac{FP}{FP + TN} = \frac{2}{1+2} = 66\%$$

● ● loan repaid men / women ○ ○ Defaulted loan men / women

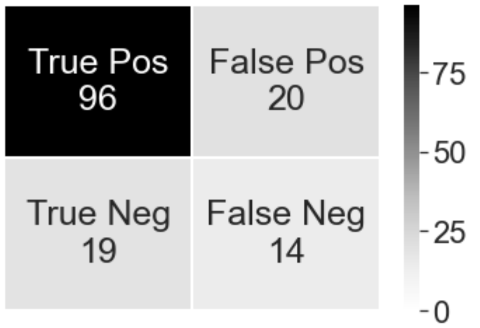
Deep dive: Comparison between the female vs. male rates in the non-adjusted model reveals unfairness across all metrics

Confusion matrix

female ♀



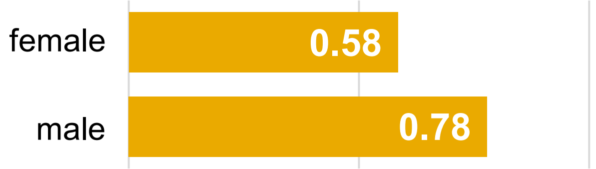
male ♂



Resulting fairness metrics

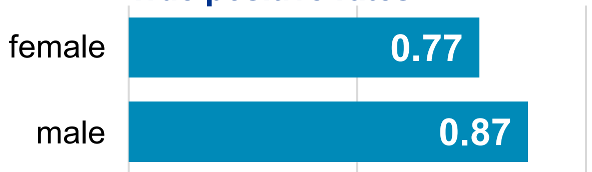
Differences in rates

Selection rates



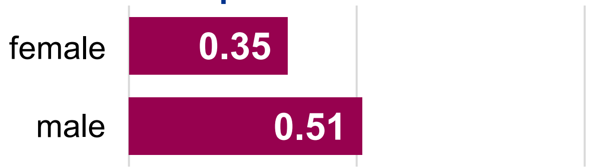
0.19 > 0.05 **unfair**

True positive rates



0.10 > 0.05 **unfair**

False positive rates



0.16 > 0.05 **unfair**

Bias can both be introduced and mitigated at every step of a model's implementation

Sources of bias

Data collection

Data is often collected by humans, which are biased themselves. In ML models usually not all data can be used. Deciding which data to chose and to neglect is a source of bias

Transparency / Explainability

The way the **model** was developed or how the model was trained results in unfair outcomes (measurement bias)

Fairness, ethics

After the model is trained, the results need to be **interpreted** for further processing of the results.

Bias mitigation

Pre-Processing

The input data can be **altered directly** in a way that corresponds to the appropriate definition of fairness, by, for example, creating synthetic data for underrepresented groups
(Re-Sampling, Relabeling, Reweighing)

In-Processing

The model should be created in a way that it satisfies certain **fairness criteria**. Constraints during training can be set, which take fairness next to accuracy into account.
(Adversial Debiasing, Prejudice Constraints, Exp Gradient Reduction)

Post-Processing

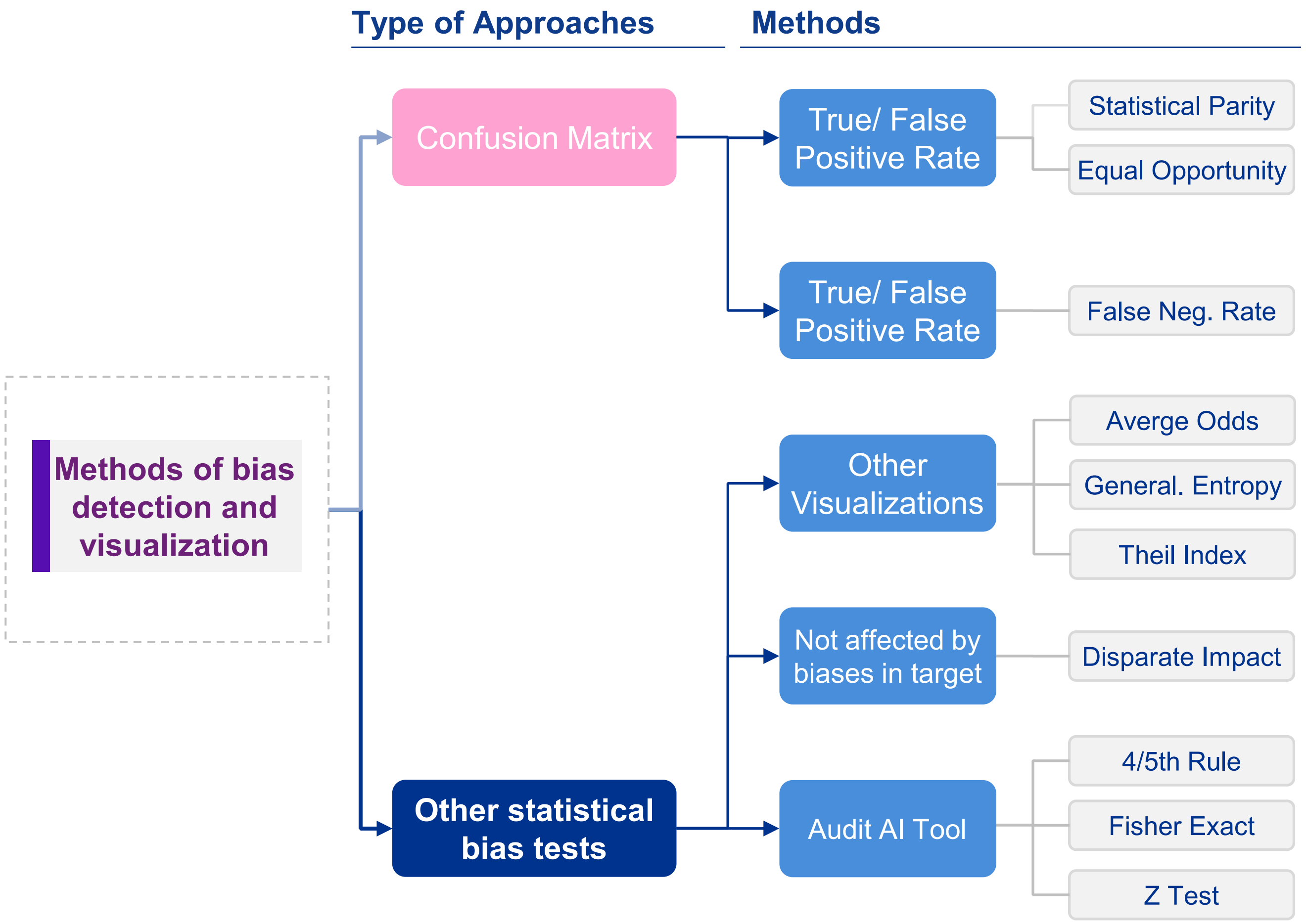
Bias can be corrected in the post-processing phase by directly **adjusting the outputs**, for example by making it easier for certain minority groups to get a "positive model outcome“
(Calibrated Equalized Odds)

As bias can be introduced at all three levels - i.e. in the data, the model and its usage – bias mitigation can and needs to use a multilevel approach as well.

In addition to approaches using the Confusion Matrix, there is a plethora of advanced statistical bias tests

Overview: methods of bias detection

- The most commonly known metrics for model performance like recall, accuracy and precision can all be derived from the confusion matrix
- The confusion matrix can also be used as a visualization of statistical bias when applying it to protected variables
- Beyond this, there have been significant developments in the field of statistical bias tests which have added to the already existing library of tests that can be used



There is a range of different bias mitigation approaches which take different mitigation measures

Approach	Selected topics
Counterfactual fairness	Swap gender in Input Data
Pre-Processing	Training ML Algorithm without gender information
In-Processing	Using constraints in training
Post-Processing	Adjust selection thresholds for each group
Accuracy	Trade-off between performance and bias mitigation

Examples

The methods shown are only examples meant for illustration.

There are numerous other methods for mitigating bias.

Analysis with swapped genders - manipulated input data detects importance of gender in the learning process

Approach

Sex is identified as the sensitive feature. The feature values are changed from female to male and vice versa.

	Age	Sex	Housing	Saving account	Credit amount	Duration	Purpose	Risk
John	67	Male	Own	Little	1.169	6	Radio/ TV	Good
Anna	32	Female	Own	Moderate	13.832	48	Car	Bad

	Age	Sex	Housing	Saving account	Credit amount	Duration	Purpose	Risk
John	67	Female	Own	Little	1.169	6	Radio/ TV	Good
Anna	32	Male	Own	Moderate	13.832	48	Car	Bad

Properties

Advantages

- The analysis directly shows the impact of the gender on the trained algorithm
- No retraining for the analysis required
- Direct explanation compared with e.g. correlation analysis for all features

Disadvantages

- Manipulation of input data may cause artifacts and is just for analytical insights

Other analytical approaches

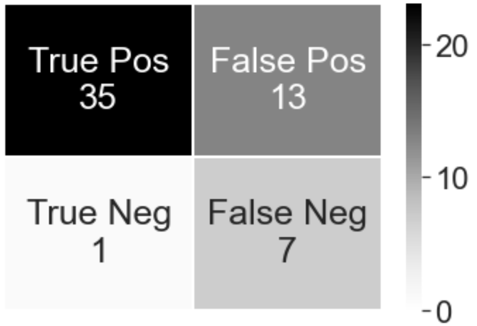
Methods

- Explainable AI (shap, LIME)
- Correlation analysis

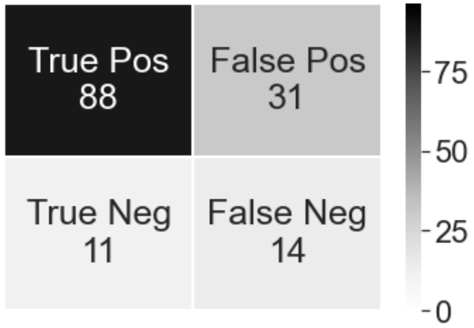
Deep dive: Analysis with swapped genders
Result: Direct, significant impact on the fairness metrics

Confusion matrix

female ♀



male ♂



Resulting Fairness Metrics

Selection rates

True positive rates

False positive rates

Differences in rates

-0.03 > -0.05 **fair**

0.02 < 0.05 **fair**

0.19 > 0.05 **unfair**

Pre-Processing: Training without gender – Manipulation of input data – In this case by removing the protected variable

Approach

Sex is identified as the sensitive feature. The feature is removed entirely from the data set

	Age	Sex	Housing	Saving account	Credit amount	Duration	Purpose	Risk
John	67	Male	Own	Little	1.169	6	Radio/ TV	Good
Anna	32	Female	Own	Moderate	13.832	48	Car	Bad

	Age		Housing	Saving account	Credit amount	Duration	Purpose	Risk
John	67		Own	Little	1.169	6	Radio/ TV	Good
Anna	32		Own	Moderate	13.832	48	Car	Good

Properties

Advantages

- The training does not take the sensitive feature into account

Disadvantages

- Other columns might correlate with the sensitive feature
- As the sensitive feature is lost, fairness cannot be controlled anymore
- Loss in accuracy

Other Pre-Processing Approaches

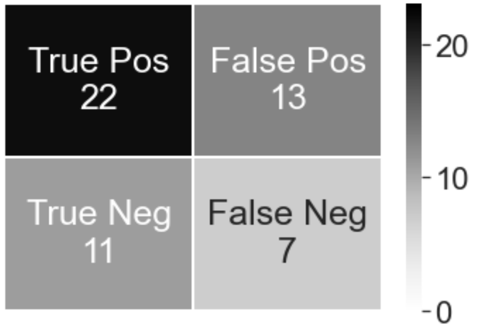
Methods

- Reweighing, Relabeling
- Synthetic Data

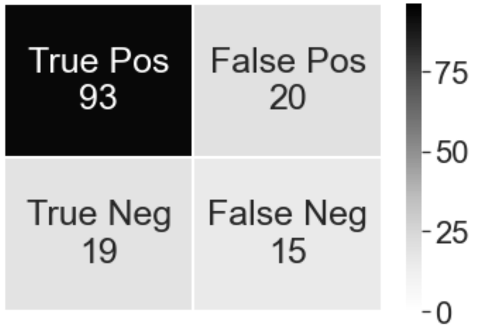
Deep dive: Training with no gender (Pre-Processing)
Result: no significant change in fairness metrics

Confusion matrix

female ♀



male ♂

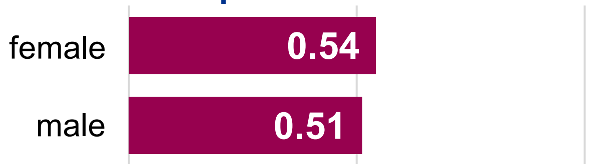


Resulting Fairness Metrics

Selection rates

True positive rates

False positive rates



Differences in rates

0.11 > 0.05 **unfair**

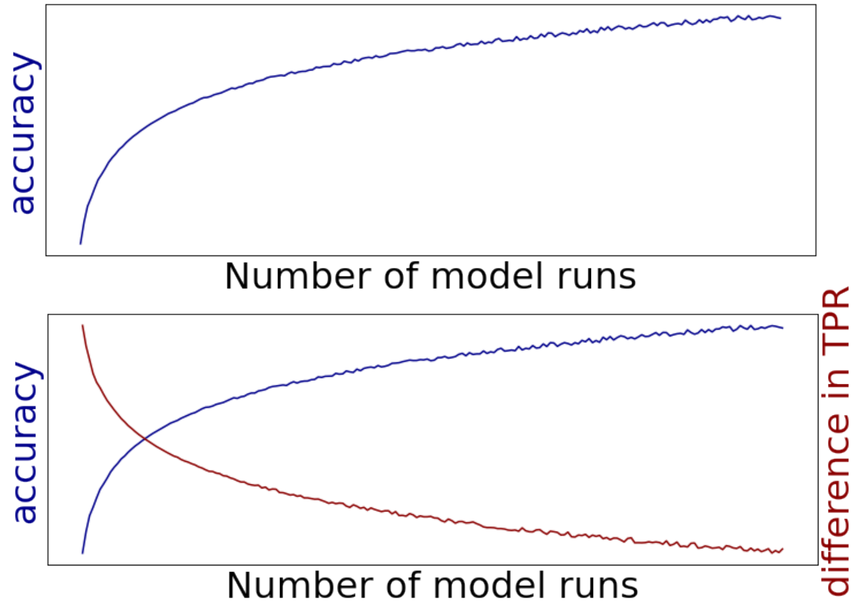
0.10 > 0.05 **unfair**

-0.03 > -0.05 **fair**

In-Processing: Adjusting the loss function taking into account constraints, for example difference in true positive rate

Approach

During model training, the model can be trained by considering not only the accuracy, but also other measures, such as the difference in the true positive rate



Properties

Advantages

- Constraints can be set directly during training
- No adjustment of input data necessary

Disadvantages

- Trade-off between accuracy and the constraint measure
- Model needs to be retrained, also when new data becomes available

Other In-Processing approaches

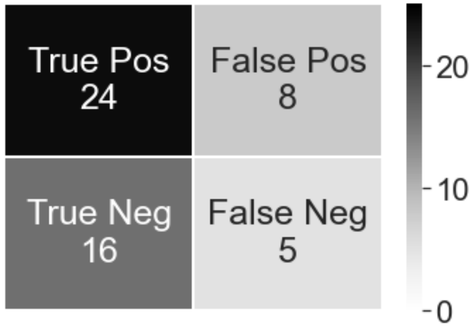
Methods

- Adversarial Debiasing
- Prejudice Remover

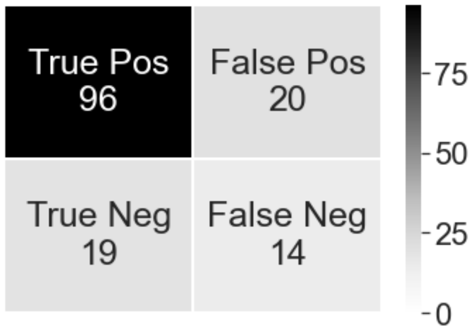
Deep dive: Constraints during training (In-Processing)
Difference in true positive rate can be minimized

Confusion matrix

female ♀

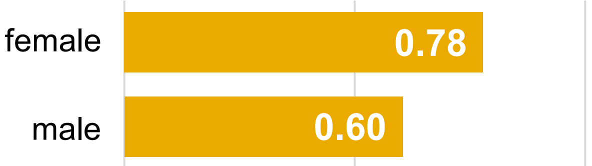


male ♂



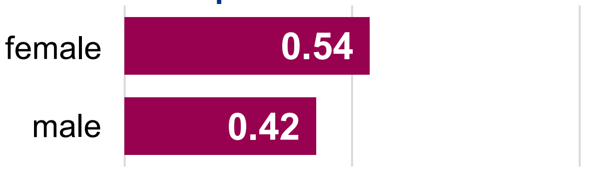
Resulting Fairness Metrics

Selection rates



True positive rates

False positive rates



Differences in rates

0.17 > 0.05 **unfair**

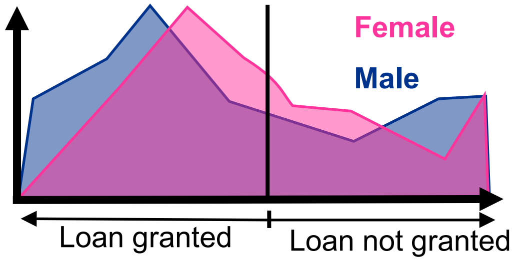
0.02 < 0.05 **fair**

0.12 > 0.05 **unfair**

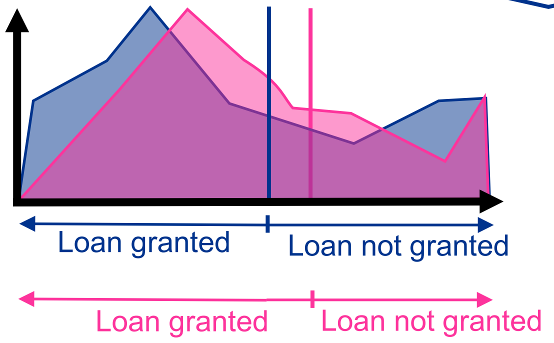
Post-Processing: Using different thresholds for a positive model outcome for female and male

Approach

The model predicts default probabilities between 0 and 1. When the probability is above 0.5, no loan is granted. The threshold can be adjusted directly for each sensitive group



	Default probability	Loan Granted?
John	0.45	Yes
Anna	0.52	No



	Default probability	Loan Granted?
John	0.45	Yes
Anna	0.52	Yes

Properties

Advantages

- No adjustments to the model or the input data necessary
- Adjustments of thresholds is effortless
- No retraining necessary

Disadvantages

- Setting of thresholds might be arbitrary

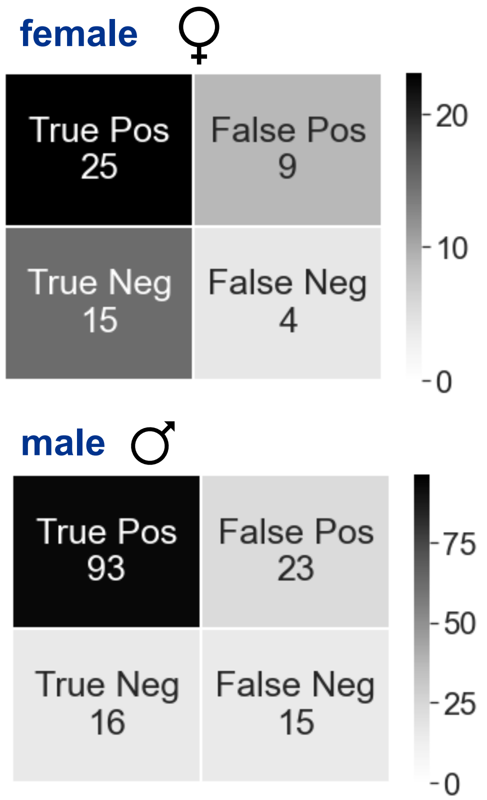
Other Pre-Processing Approaches

Methods

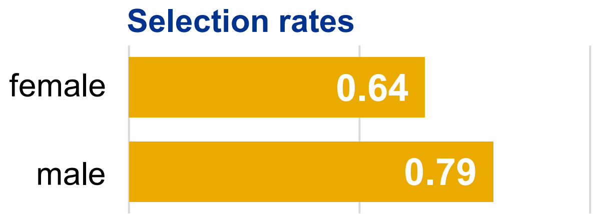
- Equalized Odds, Calibrated equalized Odds
- Classifying reject options

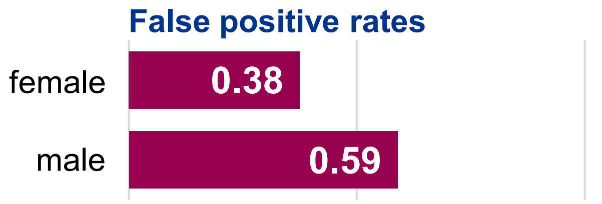
Deep dive: Adjust threshold (In-Processing) Difference in true positive rate can be minimized

Confusion matrix



Resulting Fairness Metrics





Differences in rates

0.15 > 0.05 **unfair**

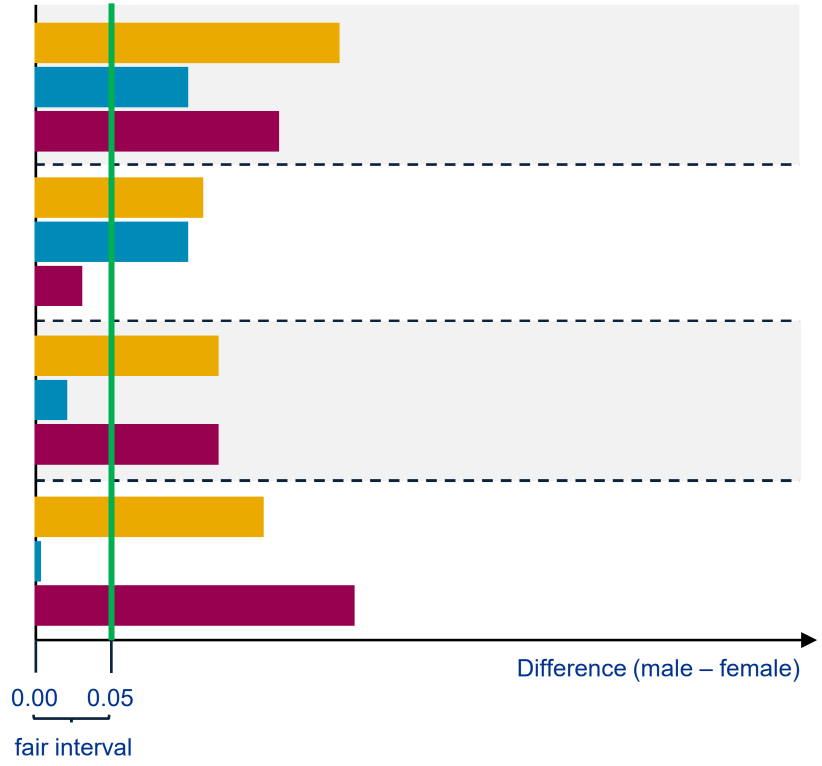
0.00 < 0.05 **fair**

0.21 > 0.05 **unfair**

Comparing model fairness after using different processing methods - not all fairness metrics satisfied simultaneously

Mitigation method	Selection rate	True positive rate	False positive rate
Original Model (before mitigation)	0.19	0.11	0.16
Pre-processing: adjust input	0.11	0.10	0.03
In-processing: train with constraints	0.17	0.02	0.12
Post-processing: adjust thresholds	0.15	0.00	0.21

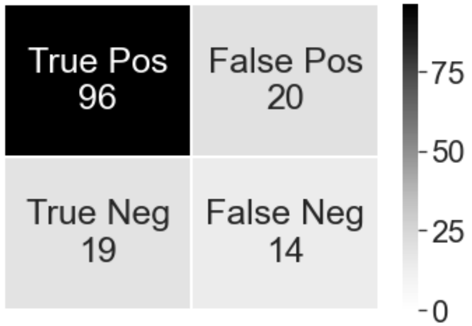
Rates per mitigation method



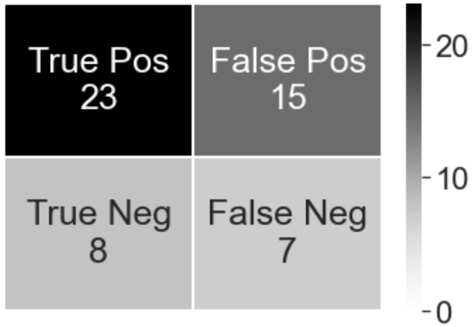
Fairness accuracy trade-off: Fairness comes at the cost of model accuracy

Non adjusted model

male – accuracy = 0.77



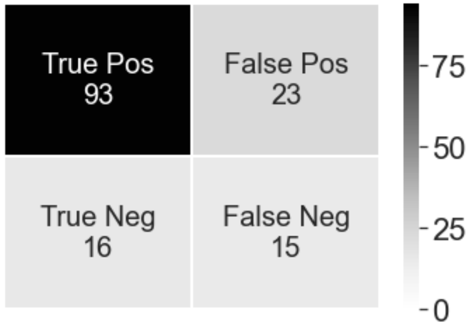
female – accuracy = 0.72



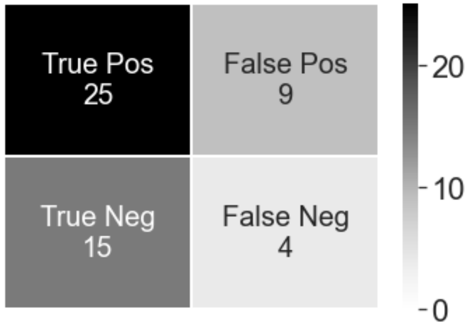
Overall accuracy = 0.76

Adjusted model (Post-Processing: adjust thresholds)

male – accuracy = 0.74



female – accuracy = 0.75



Overall accuracy = 0.74

Tradeoff

- A regular model machine learning model minimises the loss and maximises an **accuracy score**

Formula

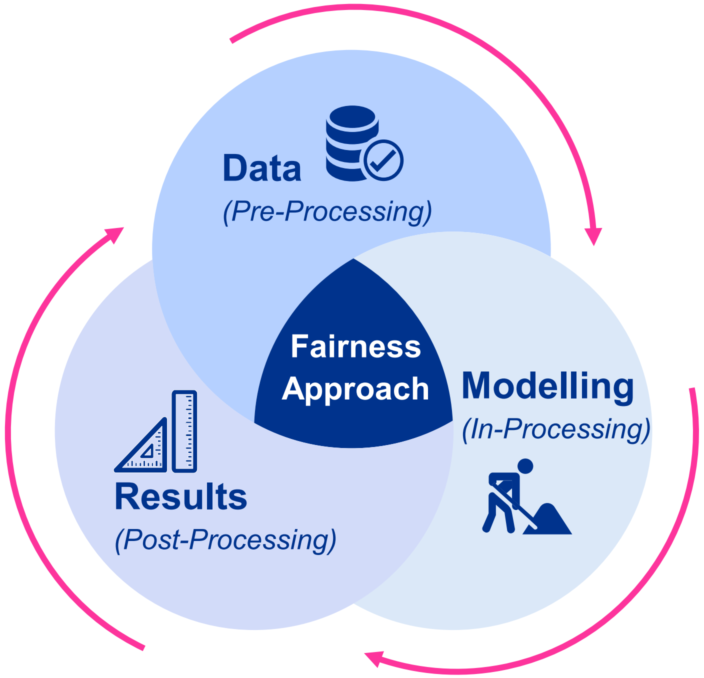
$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- A machine learning framework taking into account fairness as well, needs to find a **trade-off between fairness and accuracy**
- Generally, when **fairness increases**, the accuracy can **possibly at times decrease**

The fairness of machine learning models must be ensured using various methods

Pre-Conditions
(Fairness Methods)

- Collected Data
- Fairness Definition
- Fairness Metrics Definition
- Performance Measures



Pre-Conditions
(Tools & XAI)

- XAI Methods
- Auto ML
- AI Dashboard & AI Scorecard

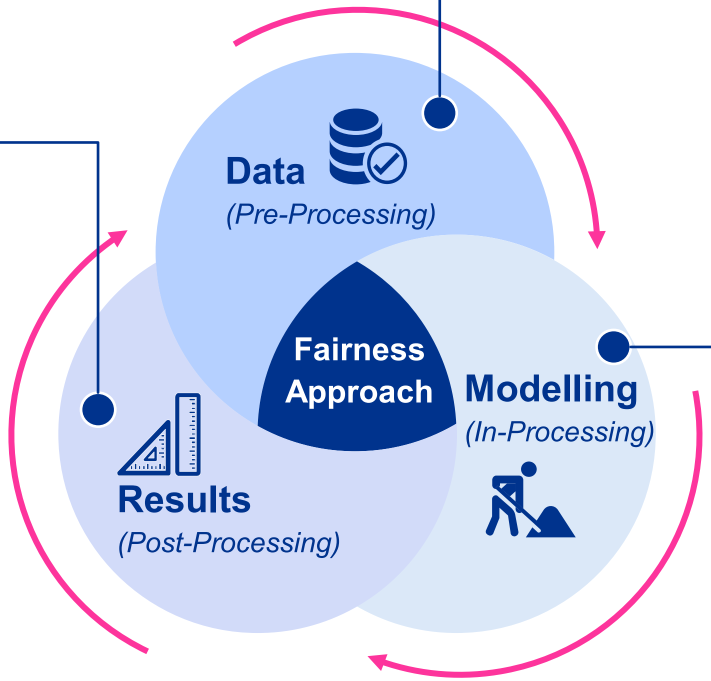
The process of achieving fairness requires a constant recursive approach, as they influence each other. In addition, after each re-training of the model, a validation of the fairness is necessary.

Different approaches are possible at each step and should be chosen according to the objectives, data & results

Post - Processing

In addition to mitigating bias via the training data or learning algorithm, post-processing algorithms can be used

- Equalized Odds
- Calibrated equalized odds
- Classifying reject options



Pre - Processing

In pre-processing algorithms, the training data can be modified to mitigate bias:

- Sampling
- Reweighing, Relabeling
- Data Transformation
- Synthetic Data

In - Processing

In-processing algorithms attempt to change the learning procedure for a machine learning model, e.g.

- Adversarial Debiasing
- Prejudice Remover
- Exponentiated-gradient reduction

Re-imagining the MRM framework for AI/ML models – Seven Key Pillars

For updating the model risk framework seven key topics that define minimal requirements for AI/ML use can be identified

1. Establish a definition of AI/ML models
2. Updating the model tiering definition
3. Establish an appropriate risk appetite
4. Identify accountability
5. Invest in skill enhancements
6. Enhance the compensatory control framework
7. Additional tests and procedures for validation

Current MRM frameworks are already very powerful tools and a great basis to handle AI / ML models. However, they need to be adjusted to fully cover the risks and challenges AI / ML models present.

For more insight into model risk management for AI models, have a look into our newest Whitepaper

More Information and Download: LINK

Executive summary

AI / ML models can offer significant added value in the delivery of financial services. However, they entrain risks that banks do not yet consider sufficiently in the existing MRM approaches.

In this paper the risks and challenges that companies encountered across the lifecycle of AI / ML models are described. In addition, seven key pillars for MRM framework AI /ML enhancements are derived.

Contact

KPMG in India

KPMG Assurance and Consulting Services LLP
Lodha Excelus, 2nd Floor, Apollo Mills Compound,
N.M. Joshi Marg, Mahalaxmi, Mumbai 400 011

KPMG in Germany

KPMG AG Wirtschaftsprüfungsgesellschaft
Klingelhöferstrasse 18 10785 Berlin

Rajosik Banerjee
Partner and Head of Financial Risk Management
rajosik@kpmg.com

Kinshuk Pal
Associate Partner, Financial Risk Management
kinshukpal@kpmg.com

Matthias Peter
Partner, Financial Services
matthiaspeter@kpmg.com

Janek Gallitschke
Senior Manager, Financial Services
jgallitschke@kpmg.com

Dr. Christoph Anders
Manager, Financial Services
christophanders@kpmg.com

www.kpmg.com/socialmedia

kpmg.com

Document Classification: KPMG Public