



Effective model risk management framework for AI/ML based models

Approaches for measuring bias and embedding fairness



October 2024

kpmg.com/in

KPMG. Make the Difference.

Foreword

In the past few years, Artificial Intelligence (AI)/Machine Learning (ML) models have become prominent instruments in a multitude of sectors. They play a pivotal role in enhancing decision-making processes, influencing business results, and increasing operational efficiency. These sophisticated algorithms have enabled unprecedented advancements in risk monitoring and assessment, fraud detection, customer service automation, and investment strategies, among other critical functions.

However, alongside these transformative benefits, the adoption of AI/ML models has introduced challenges which could potentially increase risks in the entire model lifecycle. Thus, the inherent complexity of AI/ML models amplify the importance of comprehensive Model Risk Management (MRM) frameworks to mitigate potential risks and ensure the responsible adoption and deployment of AI/ML models in financial institutions.

Effective MRM framework for AI/ML models is expected to cater to the peculiar challenges of model adoption on multiple fronts including data and model governance and accountability,

model explainability, policy frameworks, risk of bias and discrimination in the underlying data and algorithm, among other concerns. Regulatory agencies across major economies such as Singapore, United States (US) and United Kingdom (UK) are developing frameworks to tackle such challenges posed by AI/ML models. The European Union (EU) has also taken a significant stride in the direction of regulating AI through the EU Artificial Intelligence (AI) Act 2023.

This paper explores the critical necessity of an enhanced MRM framework in safeguarding against the risks associated with bias and other vulnerabilities inherent in AI/ML models. The paper also discusses the expounded checks of model validation, fairness and explainability that can be performed along with applying bias mitigation methods through a simple case study to deal with issues of model bias. By addressing these challenges proactively, financial institutions can harness the full potential of AI/ML technologies while upholding ethical standards and regulatory compliance.

Table of contents

01	Context	4
02	Different use cases of AI/ML models	6
03	Global regulatory landscape on responsible AI	8
04	Model risk management framework for AI/ML based models	12
05	Approaches for measuring bias and embedding fairness in AI/ML models	16
06	Empirical approach for model validation	20
07	Empirical case study	24
08	Conclusion	27

01

Context



Model Risk Management (MRM) for Artificial Intelligence (AI)/Machine Learning (ML) based models come with a unique set of characteristics and challenges due to the inherent complexity and dynamic nature of these models. The standard model validation framework used for conventional models requires enhancement to incorporate the potential issues such as bias, model drift, model explainability, ethical considerations, fairness-accuracy trade-offs, and diverse stakeholder engagement which are present in AI/ML based models. In this paper, we will discuss the enhancements financial institutions should make to their MRM framework to effectively manage risks throughout the entire model lifecycle. We will also discuss various techniques to measure bias and embed fairness which are key for high model performance.

Embedding fairness considerations is crucial in an AI/ML model development process. Fairness using statistical measures implies that the individuals having the same feature in every aspect, except for the value of the protected attribute (such as gender, race and marital status) should be treated equally by the algorithm. Bias, on the other hand, involves the existence of unequal or unfair treatment by the algorithm which may lead to incorrect prediction in the relationship between data

inputs and targeted output. This will impact decision-making and model predictions, resulting in an enlarged biased data for training future algorithms and could lead to regulatory and compliance issues for the financial institutions.

Thus, there exists an interlinkage between bias and fairness. When biases influence decisions, fairness is compromised due to the existence of bias in the model. Conversely, promoting fairness requires identifying and addressing biases to ensure that decisions are based on objective criteria rather than subjective preferences.

Achieving complete de-biasing of an AI/ML algorithm is simply not achievable; the objective is to reduce the presence of biases in AI/ML models. Fairness considerations should be an ongoing part of model development, model monitoring and evaluation processes.

With an increase in the development and use of AI/ML based models in recent years for different regulatory as well as key decision-making processes, in our view, it is very important that institutions strive for fairness while acknowledging and addressing inherent biases to the extent possible.

02

Different use cases of AI/ML models



The usage of AI (backed by powerful ML models) across industry applications has exploded in the last few years due to both supply-side and demand-side factors. Some of the supply-side factors include:

- Increasing computation power
- Expansion of data storage facilities
- Rise in academic research on AI/ML supplemented with the availability of data science talent
- Easy availability of vendor developed off-the-shelf AI solutions (chatbots, digital assistants, recommender engines).

The above factors have made AI adoption less costly in both financial and operational sense. AI also provides a competitive advantage to businesses which gives rise to the demand-side factors, including:

- Enabling personalised product offerings
- Better customer targeting
- Launching directed marketing campaigns
- Predicting potential lapses.

The global spending on AI based solutions is expected to increase across several industries such as finance, healthcare, manufacturing, logistics, entertainment and many more. Such growth in the usage of AI systems has led some experts and commentators to label this as the 'Fourth Industrial Revolution'.

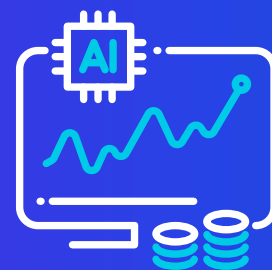
Within the financial services sector, organisations are increasingly using AI technologies in various applications such as:

- Robotic process automation for operational tasks
- Advanced AI/ML models for fraud detection
- Default prediction
- Credit lifecycles such as screening, underwriting, monitoring and collection
- Price prediction for various assets and commodities
- Macro and Micro scenario generation
- Anti Money Laundering (AML)
- Churn prediction
- Personal finance advisory
- Virtual assistants for customer services amongst other things.

Asset managers are using AI for:

- Portfolio construction
- Robo-advisory
- Risk management
- Trade execution amongst other things.

The future of financial services entails not just a rapid expansion in the portfolio of use-cases but a comprehensive infusion of AI across the organisation. While undergoing such transformation, it is pertinent to be cognisant of financial and non-financial risks arising from the use of AI/ML. One of the key elements that supersedes everything else is the ability of the model to be deemed as 'fit for purpose' which in turn has two interrelated dimensions – model fairness and model accuracy.



With an increase in usage of AI/ML models, different regulators have also started publishing different guidance notes and regulations to mitigate risk. Some of the key regulations are listed in the next section.

2. India Real Estate Vision 2047, NAREDCO and Knight Frank, August 2023

3. Real Estate Outlook 2024: Property rates projected to increase up to 15% but demand to remain steady, Mint, December 2023

03

Global regulatory landscape on responsible AI



Regulatory agencies across the world are formulating frameworks to address the unique opportunities and challenges presented by Artificial Intelligence and Data Analytics (AIDA) systems. Regulators in Asia have issued non-binding guidelines and principles on the adoption of AI, endorsing a principle based, technology neutral approach to the responsible

use of AI. Similarly, some European nations such as France, Germany and United Kingdom have come out with regulations which are mostly shaped by the European Union's data protection laws. Below table provides key regulations as published so far by different regulators:

Table 1: Key global regulations on AI and fairness

Year	Regulatory authority	Regulatory report	Key regulation summary
2024	European Union (EU) ¹	EU Artificial Intelligence (AI) Act	The EU AI Act enters into force on August 1, 2024, and will be effective from August 2, 2026. The regulations follow a 'risk-based approach' for analysing AI systems, their classification, and the formation of rules. It states that all the high-risk AI systems need to be assessed not only before being put on the market but also throughout their lifecycle.
2022	Bank of England ²	Discussion Paper: 'DP5/22 - Artificial Intelligence and Machine Learning'	The discussion paper lays emphasis on data quality with the rise in data volumes and formats within the context of AI. A need for new data quality metrics like representativeness and completeness is highlighted with the existence of bias within datasets and the AI model not performing as intended when encountering issues that are excluded from the training/testing data.
2021	Niti Aayog, India ³	'Responsible AI for All' paper	The paper aims to establish broad principles on ethics for design, development, and deployment of AI in India, leveraging similar global initiatives while adapting to the specific legal and regulatory landscape of India.
2020	European Banking Authority (EBA) ⁴	'Big Data and Advanced Analytics' report	The report outlines the concept of bias as an inclination of prejudice towards or against a person, object, or position and identifies bias detection and prevention techniques as an evolving research field.
2019	Federal Regulation ⁵	Maintaining American Leadership in Artificial Intelligence	The USA has principles for the stewardship of AI applications that prioritises fairness and non-discrimination as crucial for agencies crafting AI regulations.
2019	Financial Conduct Authority (FCA) in collaboration with the Bank of England (BoE) ⁶	'Future of Finance' report	The report aims to develop principles and share best practices for responsible uses of AI, as well as explore the intersection of AI with current rules and identify where old rules need updating.

1. Artificial intelligence act | European Parliament, July 2024

2. Discussion Paper 5/22 - Artificial Intelligence and Machine Learning | Bank of England, October 2022

3. Responsible AI for All | niti.gov.in, February 2021

4. EBA Report on Big Data and Advanced Analytics | European Banking Authority, January 2020

5. Maintaining American Leadership in Artificial Intelligence | whitehouse.gov, February 2019

6. Future of Finance: Review on the outlook for the UK financial system | bankofengland.co.uk, June 2019

Year	Regulatory authority	Regulatory report	Key regulation summary
2019	CSIRO's Data61 ⁷	Discussion paper 'Artificial Intelligence: Australia's Ethics Framework'	The paper considers fairness, one of the core principles for AI as the use of the AI systems that must not result in unfair discrimination against individuals, communities or groups and recommends a serious consideration to the degree of flexibility that designers of AI systems should have when making trade-offs between fairness measures and other priorities like profit.
2018	Monetary Authority of Singapore (MAS) ⁸	Set of 14 principles on Fairness, Ethics, Accountability and Transparency (FEAT)	The MAS has published a set of 14 principles on FEAT to encourage the deployment of AIDA in a responsible manner. To support financial institutions in implementing FEAT, MAS created Veritas consortium, offering a reliable method for integrating FEAT principles into their AIDA solutions. The Veritas is part of Singapore National AI Strategy and is a multi-phased collaborative project with financial industry. The principle of fairness emphasises two major aspects: <ol style="list-style-type: none"> 1. Justifiability wherein AIDA-driven decisions do not disadvantage any individual or groups of individuals without justification. 2. Data and models utilised for AIDA-driven decisions undergo regular review and validation to ensure accuracy and relevance, and to minimise unintentional bias.
2018	The Federal Financial Supervisory Authority (BaFin), Germany ⁹	'Big Data meets artificial intelligence' report	The report examines that there is no currently accepted standard for non-discriminating data analysis and a technical challenge exists to transform the ethical/legal definition of discrimination into a mathematical one so that it can be monitored by algorithm and prevented.

To adhere to applicable guidelines and regulations, financial institutes must enhance their model risk management framework to ensure different nuances related to AI/ML based models are assessed and monitored throughout the model life cycle.

Considering the entire landscape of issues in validating AI/ML models, this paper outlines a general framework for:

- Performing validation of an AI/ML model using alternate algorithm types
- Measuring/detecting biases in AI/ML models
- Embedding fairness by reduction/mitigation of biases present in AI/ML models
- Validation of select models post bias reduction.

7. Discussion paper 'Artificial Intelligence: Australia's Ethics Framework' | CSIRO, November 2019

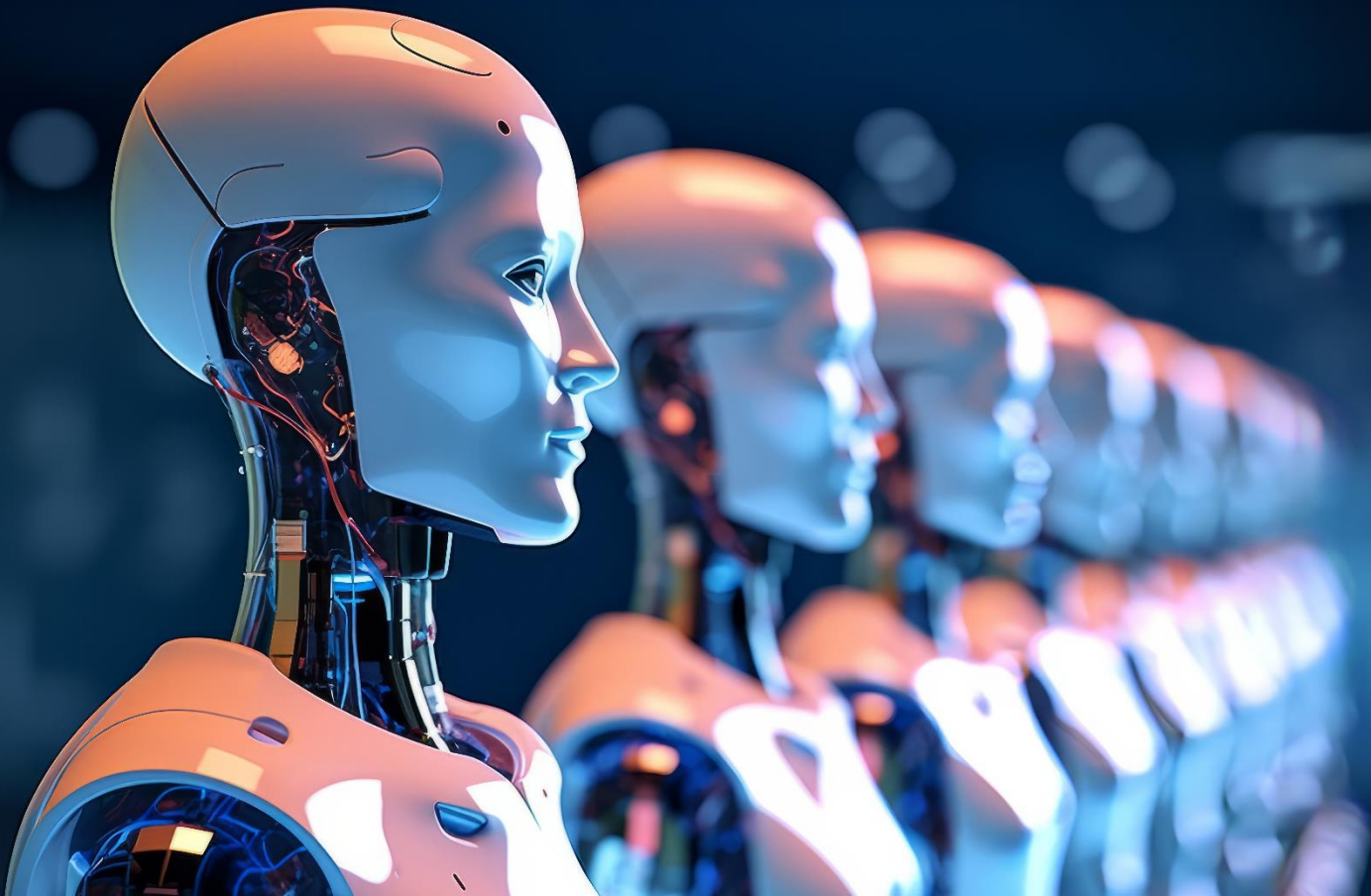
8. Principles to Promote FEAT in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector | Monetary Authority of Singapore, November 2018

9. Study 'Big Data meets artificial intelligence' | BaFin, July 2018



04

Model risk management framework for AI/ML based models



Model Risk Management framework should be enhanced for AI/ML models to address the unique challenges and complexities associated with AI/ML models. As per our view, stakeholder in three lines of defense should consider factors such as governance and accountability, complicated model architecture, model explainability, model selection, data

governance, risk of bias and discrimination, robustness and resilience, and fit for purpose assessment for such models. Such considerations would require, amongst other things, an updated quantitative and qualitative validation framework suited to identify and control the associated risks with usage of AI/ML models.

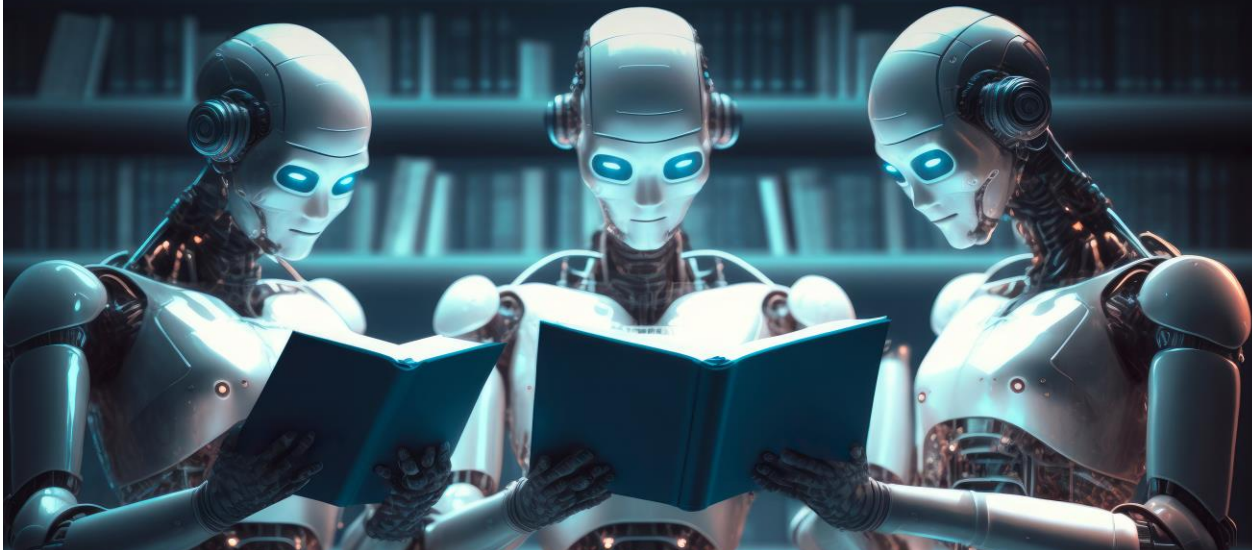
As per our view, it is precisely the variety in use-cases and methodologies in application of AI/ML models that give rise to a 'two-fold validation structure' comprising of:

1

A **model agnostic aspect** dealing with Out-of-Sample (OOS) and Out-of-Time (OOT) testing, input veracity checks, data distribution analysis, data quality checks (such as missing values and outliers) and interpretability. Typically, these aspects remain roughly constant across problems and algorithms.

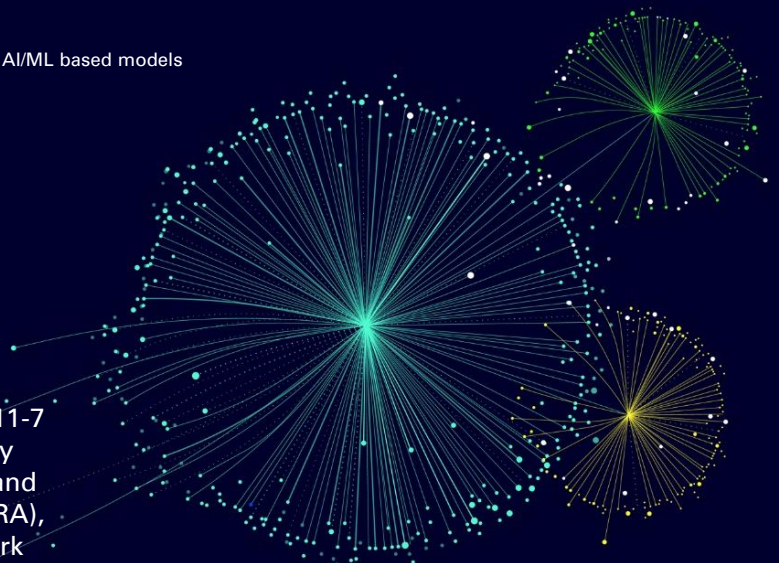
2

A **model specific aspect** like choice of algorithm, choice of target performance metric, train-test split, choice of train data, treatment of missing values, outliers, rare labels, skewed distributions and other aspects which need to be tailored to the problem at hand.



Notwithstanding the aforementioned aspects, the ethical considerations around fairness of model predictions and biases in training data give rise to an added layer of complexity. Many vendors in the market have developed tools around detection and mitigation of such ethical considerations. However, these aspects are still in nascent stages and subject to an ongoing discussion in research and practice. Choice of the right kind of detection and mitigation measures require subject matter expertise. Therefore, unchecked usage of the model may not only result in misinformed business judgement leading to potentially untoward outcomes but also in biased and unethical decision making.

Model validation framework



As per regulatory guidelines on MRM principles such as SR Letter 11-7 by Federal Reserve and Supervisory Statement SS1/23 by Bank of England Prudential Regulation Authority (PRA), a sound model validation framework should include, but not be restricted to, the following elements:

1. Defined model tiering basis complexity, usage and materiality of the model. As per our view, such models should be treated as high-risk models given inherent complexity of such models

2. Clear roles and responsibilities of all stakeholders as well as qualified and experience people in all three lines of defense for such models

3. An appropriate model validation scope and methodology should include elements such as:
 - Model inputs
 - Model theory and design
 - Model output
 - Independent model review
 - Model implementation
 - Model monitoring and control
 - Model risk reporting

4. Detailed documentation of the model validation framework and process, including documentation of the validation procedures performed, any changes in validation methodology and tools, the range of data used, validation results and any remedial actions taken where necessary

5. The findings and outcomes of model validation should be reported in a prompt and timely manner to the appropriate level of authority

6. An effective model validation process should facilitate the timely identification and resolution of potential limitations in a model

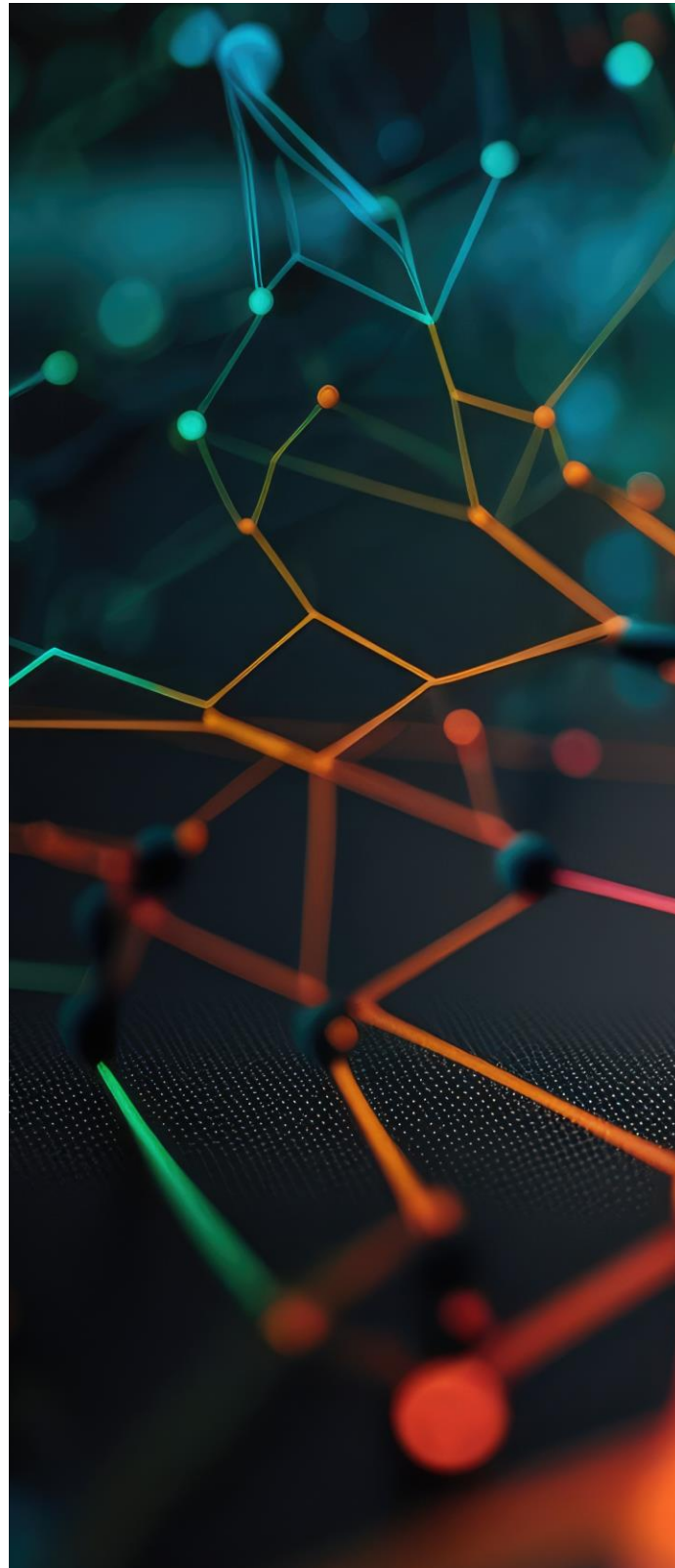
7. A review of the model validation process by independent parties (e.g., internal or external parties) to assess the overall efficacy of the model validation process and the independence of the model validation process from the development process.

AI/ML specific validation considerations

AI/ML models are used in a variety of problems within the financial services industry and are expected to become an integral part of the industry in the future. AI/ML models serve predictions as the output which are subsequently used to make business decisions, the result of which feeds back as inputs into the future training data.

As per KPMG in India's view, some additional validation that should be carried out to validate AI/ML based models are:

1. The model parameters are calibrated with the most recent training data
2. The model can detect the intended patterns in the data
3. The model doesn't overfit and generalises well to unseen datasets
4. The model hyperparameters are optimised for high performance
5. The distribution of features and target are similar in train and test sets
6. The data doesn't exhibit multivariate feature drift
7. The model results are interpretable by the users and do not come from the dreaded 'black-box'
8. The model's inference time is within acceptable range depending upon the intended application
9. The model adheres to established standards of unbiasedness towards protected groups and doesn't discriminate on the basis of gender, religion and race
10. The model effectively reduces dimensionality, especially when dealing with large feature spaces
11. On-going model monitoring, back-testing and attribution analysis for different data parameters to ensure fairness and unbiasedness in the underlying models.



05

Approaches for measuring bias and embedding fairness in AI/ML models



Fairness is commonly defined as the state of being impartial and equal without any favouritism or discrimination. However, the definition and interpretation of fairness can differ in different contexts across different fields.

Demonstrating fairness requires identification of the individuals and groups that may be subject to systematic disadvantage, determination of the harms and benefits created by the system, and measurement of the same across individuals and groups to assess systematic disadvantage. Fairness can be measured by comparing the models' predictions across groups based on protected attributes (e.g., gender, age groups, marital status, among others) that may have potential fairness considerations.

AI/ML models may result into unfair treatment in which some individuals or groups of people are privileged (i.e., receive a favourable treatment) and others are unprivileged (i.e., receive an unfavourable treatment) and decisions are based on sensitive or protected variables (such as gender, ethnicity, race, religion, disability and more). Modeling fairness in AI/ML is thus a key requirement to correct such bias in the model.

The first step in the adoption of effective bias detection mechanism is understanding the various causes of biases such as:

- Training data bias
- Algorithmic bias.

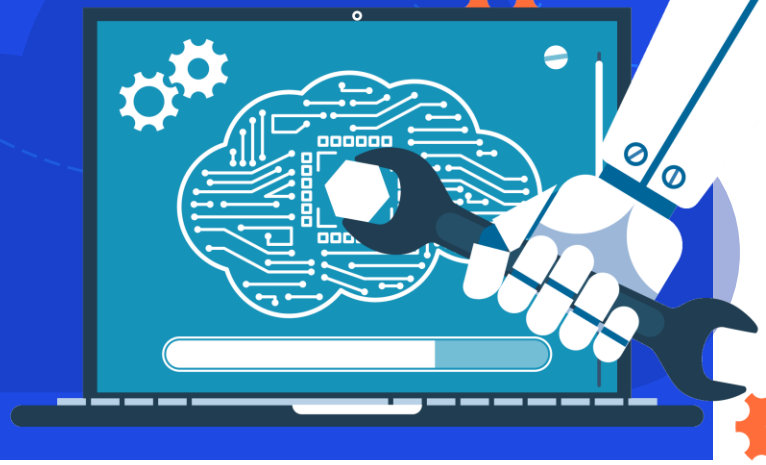
5.1 Statistical detection of model biasedness

The goal of monitoring and detecting bias is to achieve an equal probability of population groups to receive a positive treatment, or an equal treatment of individuals that only differ in sensitive/protected attributes (which partitions a

population into groups whose outcomes should have parity e.g., race, religion and gender). There are some open-source toolkits available for testing fairness of the AI/ML models.

The statistical measures of fairness are based on the following bias detection metrics:

- Statistical parity
- Equal opportunity difference
- Average odds difference
- Disparate impact
- Thiel Index
- 4/5th Rule
- Bayes Factor Test
- Counterfactual fairness
- Fisher Exact Test
- Chi Squared Test



5.2 Mitigation of model biasedness

The objective of bias mitigation in AI is to unlock value responsibly and equitably. Bias mitigation algorithms are categorised based on the stage of the AI/ML process in which they are deployed: Pre-processing, In-processing and Post-processing methods.

Pre-processing bias mitigation

- Bias mitigation algorithm is applied to training data, which is used in the first step of the AI/ML process. The types of pre-processing mitigations can range from simple data preparation methods to more complex methods like optimised data transformation which reduces bias and the predictability of the protected attribute
- Methods used are:
 - Sampling
 - Reweighting
 - Relabeling
 - Data transformation.



In-processing bias mitigation

- In-processing algorithms offer unique opportunities to reduce bias and increase fairness during the training of a machine learning model
- Methods used are:
 - Adversarial debiasing
 - Prejudice remover
 - Exponentiated-gradient reduction.



Post-processing bias mitigation

- Bias mitigation algorithm is applied to predicted labels. Such methods are applied post successful training of the classification model
- Methods used are:
 - Equalized odds
 - Calibrated equalized odds
 - Classifying reject options.



It is important to note that it is not possible to entirely mitigate bias from any AI/ML algorithm. The mitigation approaches can help to reduce the severity of the bias towards a particular class, keep the bias within limits and thus, can help institutions strive towards achieving greater fairness in their practices and outcomes.





06

Empirical approach for model validation

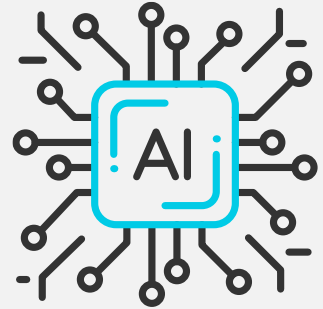


6.1 Model validation tests

Depending on the nature of the underlying dependent variable, suitable supervised learning models (classification model for categorical and regression model for continuous dependent variables) can be applied to train the data. The metrics used for Out of sample performance testing are standard in the Machine Learning community and industry.

Out of Sample performance testing metrics:

- Accuracy
- Area under the ROC Curve
- Gini Coefficient
- Precision
- Recall
- F1 score



Validation of AI/ML models is one of the most important aspects of developing any model. Validation can be performed through various tests and checks. **Deepchecks** is an open-source python package which uses a collective process known as **Suite** to validate AI/ML models and the underlying data, which involves various types of checks along with the passing/failing status for each tests conducted.

The suite includes checks like train test performance, feature label correlation, multivariate drift, feature drift, label drift, datasets size comparison among others. These checks fall into 4 major categories:

Data distribution

to check the **similarity** between test data and training data

Data integrity

to check **accuracy** of the underlying data used in the model

Methodology

to check whether **train/test sets sized correctly** and **free of leakage**

Performance

to check the **performance** of underlying model

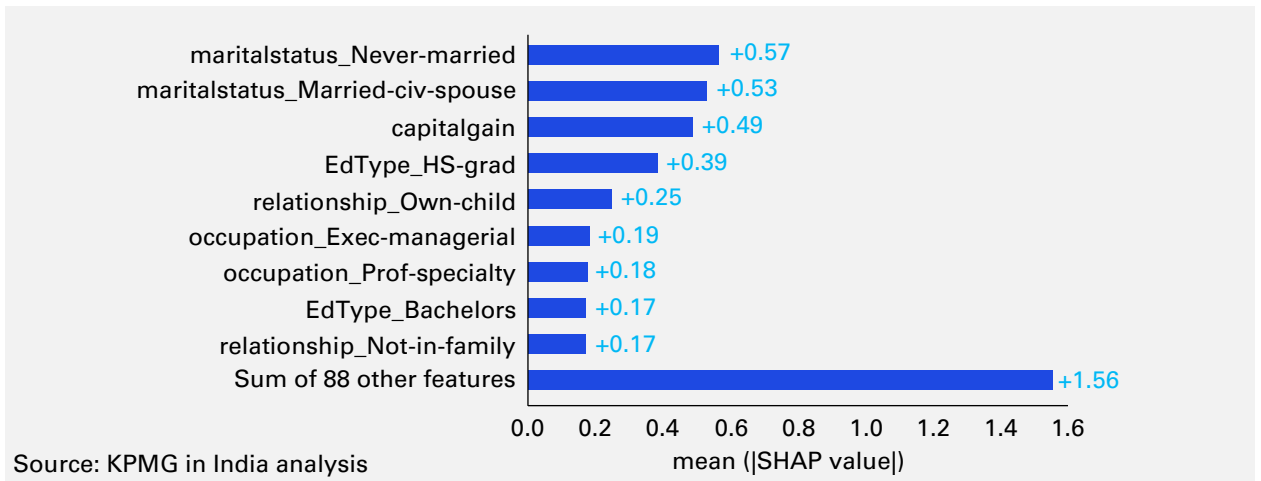
Issues like data integrity and data drift can sometimes go unnoticed while validating different AI/ML models. Hence, this platform serves as a powerful validation tool and runs several validation checks across all the above categories.

6.2 Model explainability

To understand the results produced by the different statistical models, Shapley Values can be used which is a well-known model agnostic Explainable AI technique used to understand predictions from a model in terms of the underlying features. In particular, by computing the contribution of each feature in final outcome classification, one can get a better sense of the ‘black-box’ predictions from complex AI/ML models.

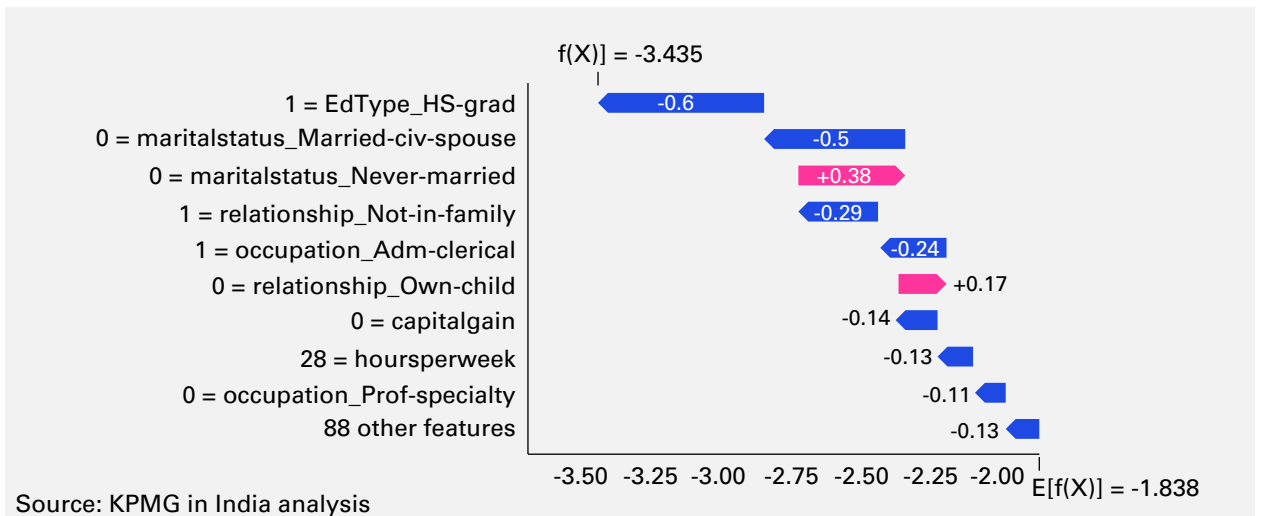
Following is an illustrative feature importance plot which demonstrates the importance of different features on the model’s predictions in a decreasing order. For each feature, the mean of absolute SHAP value is computed across all observations. Larger mean SHAP value indicates larger positive/negative contribution made by a particular feature in outcome prediction.

Illustrative Feature Importance Plot



The SHAP value of each feature can also be visualised for each observation. An indicative waterfall plot for one observation is depicted below. The SHAP values demonstrate how the features have contributed to the prediction when compared to the mean prediction. Large positive/negative values indicate that the feature had a significant impact on the model’s prediction.

Illustrative Waterfall Plot



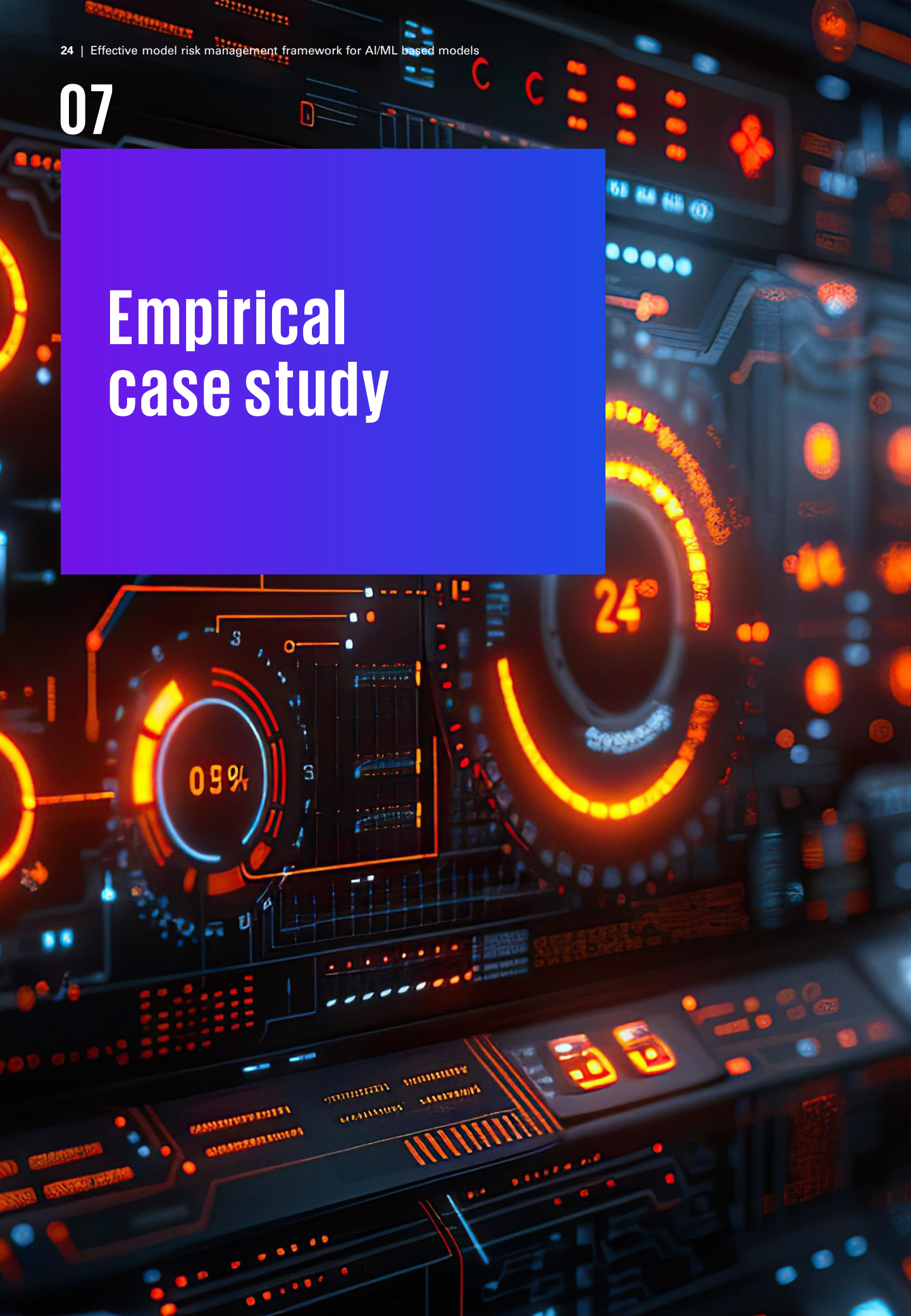
6.3 Comparison of pre-and post-bias mitigation models

- A protected attribute of interest for a particular dataset should be identified in the first step before proceeding for bias detection and mitigation
- Next, the binary dependent variable of the underlying dataset should be categorised as a favourable outcome and an unfavourable outcome. The class or group having a higher proportion of favourable outcome as compared to that of the other class or group is considered to be the privileged group and the other group with lower proportion of favourable outcome is considered to be the unprivileged group
- Next, different bias detection algorithms can be applied to demonstrate the existence of biases in the corresponding protected attribute.
- Post bias detection, pre-processing, in-processing and post-processing bias mitigation techniques can be applied for the purpose of mitigating bias in favour of the privileged group. Post bias mitigation, it is expected that both the privileged and unprivileged group should have almost same proportion of favourable outcome, indicating bias has reduced
- However, there exists a trade-off between fairness and accuracy of the model. As the fairness is achieved in the model by reducing bias, the predictive accuracy of the model can be compromised to some extent.



07

Empirical case study



In this section, a simple case study is presented where an open-source income classification dataset is considered which contains data of 31,978 customers with different attributes like age, job type, marital status, occupation and race. The dependent variable is 'Salary Status' and objective is to predict salary status of each customer (with 2 categories: Salary \leq 50,000 and Salary $>$ 50,000).

Step 1

The binary dependent variable of the dataset has been classified as 0 for Salary \leq 50,000 (considered as unfavourable outcome) and 1 for Salary $>$ 50,000 (considered as favourable outcome) respectively. The categorical attributes (such as marital status, job type and race) are converted into numerical indicator variables. The protected attribute for the particular dataset has been considered as 'Gender'.

Step 2

Following the data transformation and data cleaning phase, the entire dataset is divided into training and test dataset with 85 per cent and 15 per cent train-test split.

Step 3

Supervised learning algorithms such as Logistic Regression, Random Forest and Adaboost are built on the training dataset.

Step 4

Out-of-Sample performance testing (including precision, recall, F1-score, accuracy) of each fitted model is performed in both training and test dataset. Also, model validation tests are performed using Deepchecks validation suite to analyse the underlying data and model.

Step 5

Model explainability tests are performed to understand which features are affecting the model's predictions the most using the explainable AI tool- SHAP.

Step 6

Next, biases are explored in the protected attribute of the dataset (i.e., 'Gender') using different bias detection metrics, such as statistical parity, equal opportunity difference and average odds difference among others.

Step 7

Post bias detection, an in-processing bias mitigation technique 'Exponentiated Gradient Reduction' is performed for debiasing the dataset.

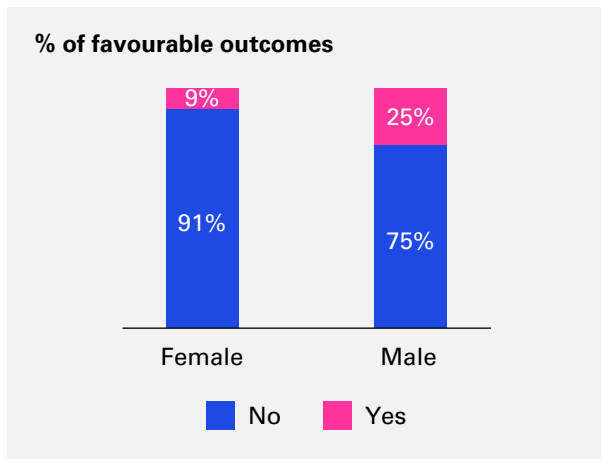
Step 8

In the final step, the bias metrics are recomputed post bias mitigation and finally a comparison of bias and accuracy of the models pre and post applying bias mitigation technique is performed, which demonstrated the trade-off between fairness and accuracy of the models.

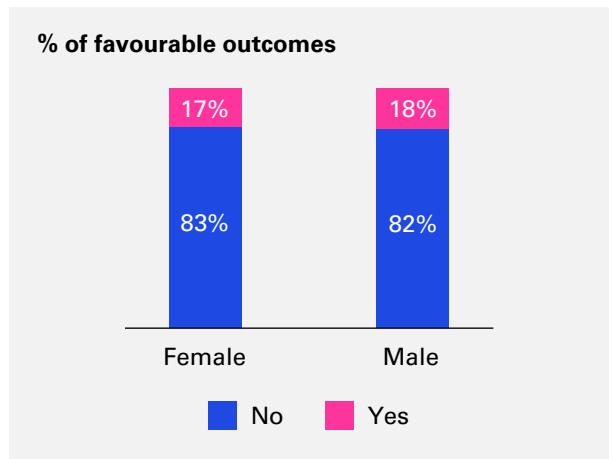


For the purpose of demonstration, a graphical representation of comparison of bias pre and post bias mitigation is depicted below for logistic regression model. From the below graph, it can be seen that, before bias mitigation technique has been applied, 'Male' category has higher per cent of favourable outcome (25 per cent) as compared to that of 'Female' category (9 per cent), which indicates that 'Male' is the privileged group and 'Female' is the unprivileged group. Post bias mitigation, 'Male' category and 'Female' category have almost same per cent of favourable outcome (with 18 per cent and 17 per cent of favourable outcome respectively), indicating bias has reduced in the model post applying the bias mitigation technique.

Pre-Bias Mitigation



Post-Bias Mitigation

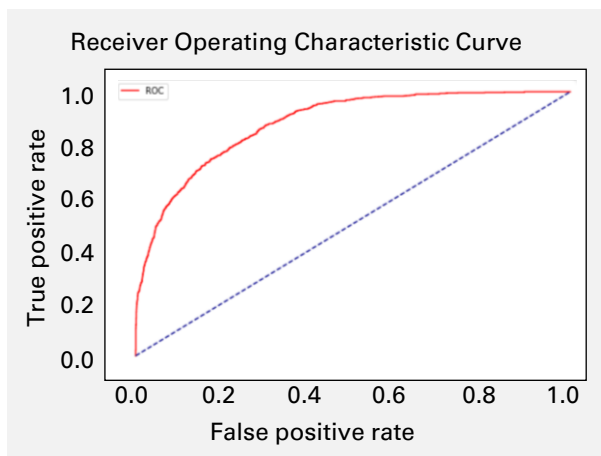


Source: KPMG in India analysis

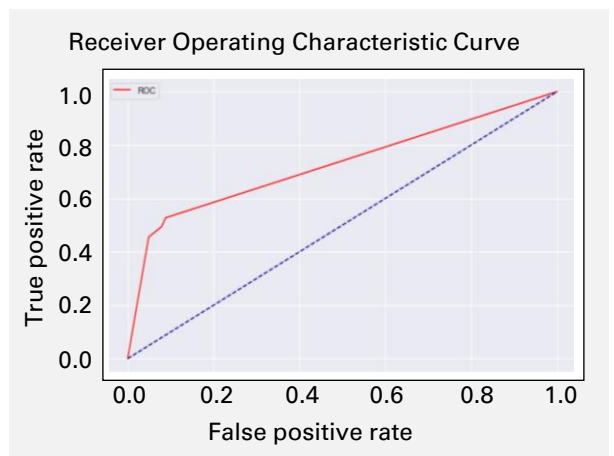
Post applying bias mitigation technique, model predictive accuracy has reduced slightly from 84 per cent to 82 per cent in test set and from 84 per cent to 83 per cent in training set.

Following is the comparison of ROC curves – Pre and Post-bias mitigation technique has been applied. As it can be seen from the below graphs, the area between ROC curve and diagonal line has reduced post applying bias mitigation, which demonstrates that there always exists a trade-off between fairness and accuracy.

Pre-Bias Mitigation



Post-Bias Mitigation



Source: KPMG in India analysis

Conclusion

With the rapid expansion in the use of AI/ML models, it is essential for institutions to enhance their MRM framework to ensure effective monitoring and control of the associated risks. The existence of bias in the AI/ML models can pose several challenges that can adversely affect the decision-making process of these institutions using such models.

While this paper provides tools for bias detection, bias mitigation and model explainability in the context of a model lifecycle, it is important to keep in mind that the notions of bias and fairness are mostly application driven or context sensitive; in other words, the choices of the attributes for measuring bias, as well as the choice of the bias metrics, can be guided by legal, social, and other non-technical considerations. The successful adoption of fairness-aware AI/ML approaches requires a thorough understanding of the characteristics of the AI/ML models in use, as well as the appropriate bias detection and mitigation algorithms. Achieving this also involves fostering collaboration across key stakeholders including AI/ML teams and end users of the models.

Acknowledgments

Financial Risk Management (FRM) Team

Rachit Gupta - Director
Riddhima Sobti - Assistant Manager

Design Team

Shveta Pednekar

Marketing Compliance

Pooja Patel



KPMG in India contacts:

Akhilesh Tuteja

Head – Clients & Markets

E: atuteja@kpmg.com

Manoj Kumar Vijai

Office Managing Partner-Mumbai

Partner and Head – Risk Advisory

E: mkumar@kpmg.com

Rajosik Banerjee

Deputy Head – Risk Advisory

Partner and Head – Financial Risk Management

E: rajosik@kpmg.com

Amitava Mukherjee

Partner – Financial Risk Management

E: amitava@kpmg.com

Somdeb Sengupta

Partner – Financial Risk Management

E: somdebsengupta@kpmg.com

kpmg.com/in



Access our latest insights
on KPMG Insights Edge

Follow us on:

kpmg.com/in/socialmedia



The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation.

KPMG Assurance and Consulting Services LLP, Lodha Excelus, Apollo Mills Compound, NM Joshi Marg, Mahalaxmi, Mumbai - 400 011 Phone: +91 22 3989 6000, Fax: +91 22 3983 6000.

© 2024 KPMG Assurance and Consulting Services LLP, an Indian Limited Liability Partnership and a member firm of the KPMG global organisation of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organisation.

This document is for e-communication only. (011_THL0924_SP)