Reliable and fair machine learning for risk assessment

Pereira Barata, A.P.

Reliable and Fair Machine Learning for Risk Assessment

António Pereira Barata

Human Environment and Transport Inspectorate
Ministry of Infrastructure and Water Management

Universiteit Leiden
The Netherlands

The work in this book was funded by the Ministry of Infrastructure and Water Management through Leiden University. The author used facilities at the Leiden Centre of Data Science, the Leiden Institute of Advanced Computer Science, and the Innovation and Data Lab of the Human Environment and Transport Inspectorate.

SIKS Dissertation Series No. 2023-06
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Typeset using LaTeX, figures generated using `Matplotlib`
Printed by: Gildeprint, Enschede, the Netherlands, gildeprint.nl
ISBN: 978-94-6419-719-8

Reliable and Fair Machine Learning for Risk Assessment

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 5 april 2023
klokke 13:45 uur

door

António Pereira Barata
geboren te Lisboa, Portugal
in 1989

Promotor:	Prof.dr. H. J. van den Herik	
Co-promotores:	Dr. C.J. Veenman Dr. F.W. Takes	
Promotiecommissie:	Prof.dr. T.H.W. Bäck	(TU Delft)
	Prof.dr. C.M. Jonker	(Umeå University)
	Prof.dr. M.V. Dignum	(Universiteit Twente)
	Prof.dr. J.N. Kok	(RWTH Aachen University)
	Prof.dr. H.H. Hoos	

Faculteit Wiskunde en Natuurwetenschappen

Dedicated to Alexa

"It is not your fault, but it is your responsibility."

Cor J. Veenman

Contents

List of Abbreviations

AI	artificial intelligence
AP	average precision
AUC	area under the receiver operating characteristic curve
CV	cross-validation
EU	European Union
FAHT	fairness-aware Hoeffding tree
FPR	false positive rate
IF	isolation forest
ILT	Inspectie Leefomgeving en Transport
KNNC	k-nearest neighbours classifier
LOF	local outlier factor
LR	logistic regression
MAR	missing at random
MCAR	missing completely at random
MNAR	missing not at random
NAR	noise at random
NCAR	noise completely at random
NNAR	noise not at random
NWF	non-white female
NWM	non-white male
OvR	one-versus-rest
PS	problem statement
RQ	research question
SCAFF	splitting criterion AUC for fairness
SVMC	support vector machine classifier
TPR	true positive rate
WF	white female
WM	white male
XGBC	extreme gradient boosting classifier
XGBR	extreme gradient boosting regression

List of Definitions

List of Figures

List of Tables