The double-edged sword of Big Data and AI for the disadvantaged: a cautionary tale from Open Banking

by Savina Kim, Galina Andreeva, Michael Rovatsos

Galina.Andreeva@ed.ac.uk

Working Paper

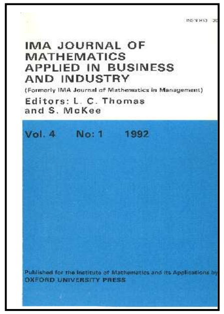
CRC Credit Research Centre

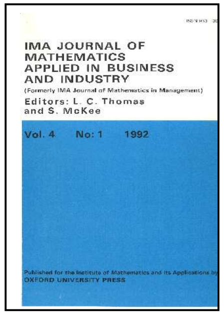
The **Credit Research Centre (CRC)** is an impartial independent research group devoted to the study of credit

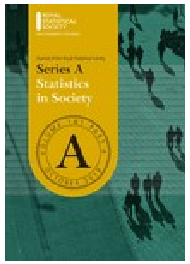
- Founded in **1997**
- Successful collaboration with the industry and regulations since **1989** (1st Credit Scoring & Credit Control conference)
- **20th** conference - Edinburgh 31 August to 03 September 2027
- Seminar and webinar series, Village Hall discussions, working papers, and much more.

The last conference was attended by 430 delegates from162 organisations (Universities, financial institutions, Central Banks) and from 41 countries

13 Special Issues + 5 Impact Case Studies







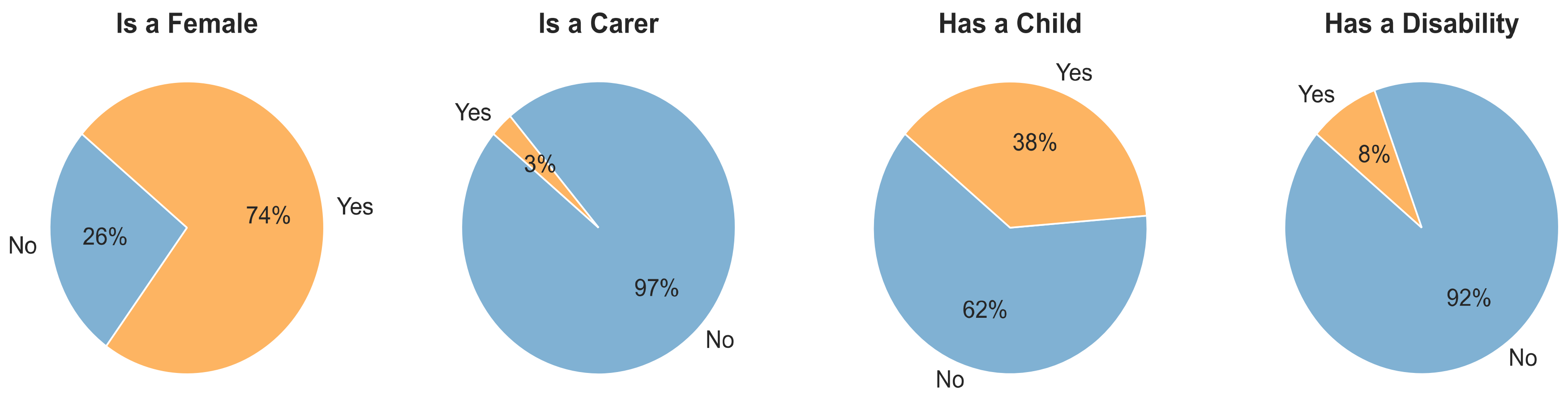
Summary

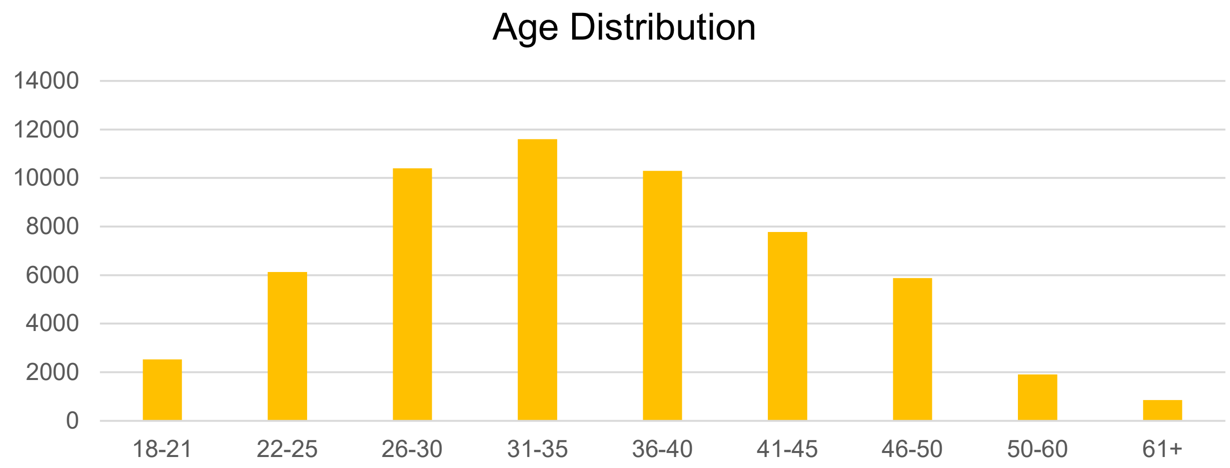
- Motivation: expansion of Big Data and AI (although the results are also relevant for small datasets and traditional models)
- Hidden dangers, e.g. amplified algorithmic bias, specifically ‘proxy’ bias
- Open Banking (OB) as an example of Big Data
- Data description
- Regulatory context, protected and sensitive attributes
- Power of OB data: Financial Vulnerability (FV) profiling
  - Indicators and specification
  - What if someone wants to target a vulnerable group? Predicting FV – machine-learning wins
- Associations with protected and sensitive attributes:
  - What it means for the probability of being accepted for credit
  - What it means for customer segmentation (not presented here)
- Conclusions and recommendations: It is essential to recognise and address ‘proxy’ bias in data to prevent inadvertently disadvantaging protected groups.

Open Banking/ transactional data

- The second Payment Services Directive (PSD2 or Open Banking in the UK) is a game changer in retail finance. It enables easier access to dynamic, real-time consumer data in practice, with bank transaction data being of particular interest.
- The research uses a new, proprietary dataset provided by a UK social lender, offering small loans (£500 to £1,000) to financially challenged public sector workers. The lender assesses applicants via Open Banking during the affordability check procedure, excluding the need for a bureau credit score.
- The data was collected in February 2022 and focuses on approximately 100,000 applicants who have applied for a loan in the previous two years, yielding a dataset of over 180 million transactions.

Protected and sensitive attributes





Legal position

The Law makes the distinction between direct and indirect discrimination.

It is concerned with procedural fairness or '**equal treatment**' or 'direct discrimination' which is strictly prohibited → certain variables cannot be used as inputs into a model (be it a regression or a machine-learning algorithm).

The public and media seem to be more concerned about outcome fairness or '**equal outcome**'. However, unequal outcomes can arise from 'indirect discrimination', where an apparently neutral criterion would put persons of a particular group (e.g. gender) at a disadvantage compared with other persons. This can be justified by a legitimate aim, given the means of achieving that aim are appropriate and necessary.

Equal treatment does not automatically result in equal outcome, see Andreeva, G., Matuszyk, A.: The law of equal opportunities or unintended consequences: The impact of unisex risk assessment in consumer credit. Journal of Royal Statistical Society, Series A, 182(4), 1287—1311 (2019)

Financial Vulnerability modelling

- Financial Vulnerability (FV) modelling follows regulatory guidance –FCA Financial Lives survey
- Proposed six binary FV indicators (target variables), e.g.:
  - *Financial shock withstanding* (48.3% of applicants): average median monthly balance below £100.
  - *Returned direct debits* (28.3%): When an applicant has at least one or more returned direct debits (RDD), insufficient funds resulting in a rejection of a pre-arranged payment by the bank.
- Over 100 predictor variables, e.g. :
  - Financial Distress (Debt management payments, overdraft, RDD)
  - Financial Inclusion (# credit providers, # cards, loans, payday loans).
- Compared Logistic regression, Random Forest, eXtreme Gradient Boosting
- Very high predictive accuracy in modelling FV (AUC 0.756-0.911) and sensitive attributes (AUC 0.796-0.917), i.e. easy to profile and target.
- Observed substantial disparities for protected groups in simulated scenarios.

Financial Vulnerability indicators/ binary target variables

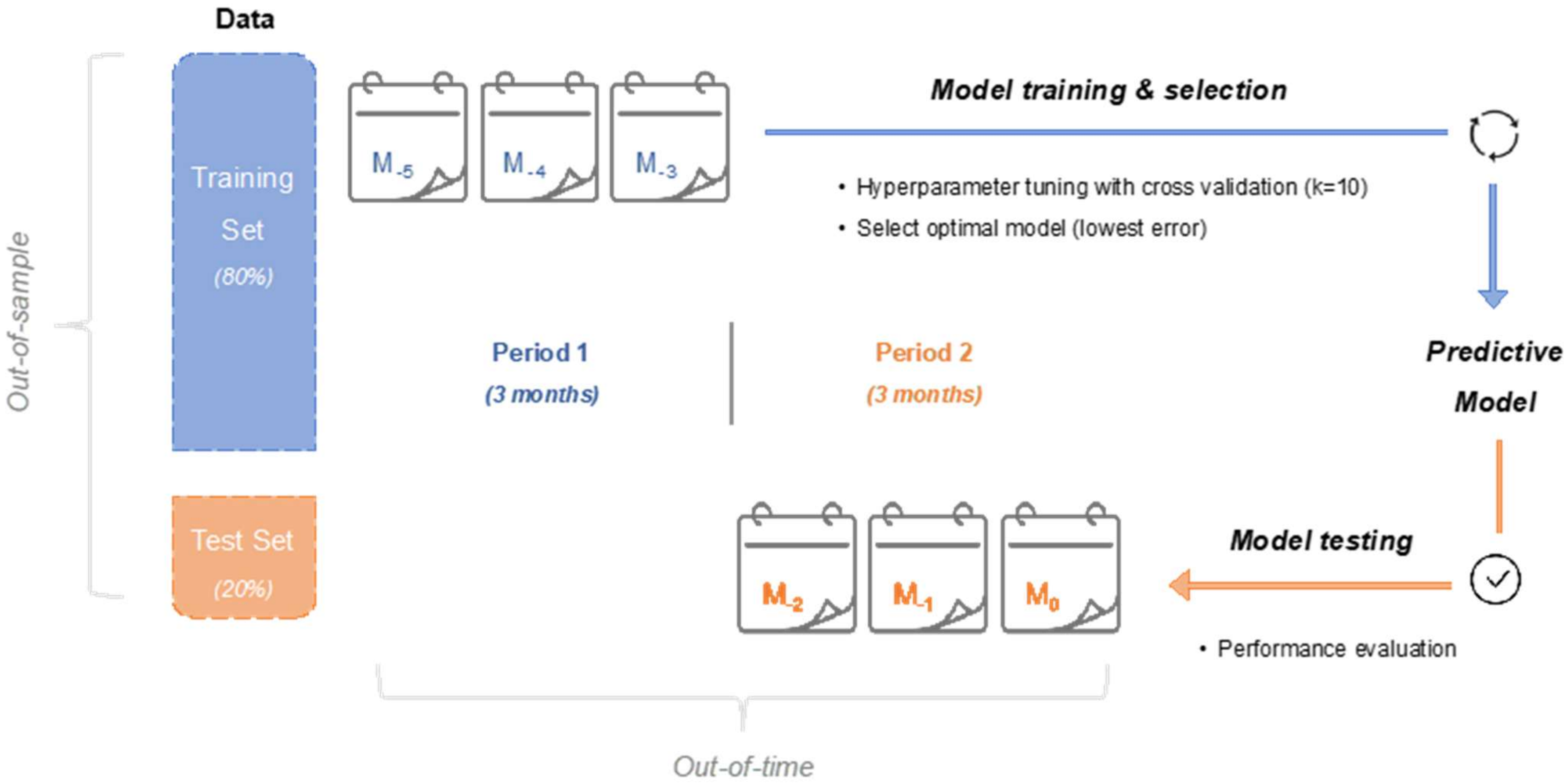
1. *Financial shock withstanding* (48.3% of applicants): average median monthly balance below £100.
2. *Insolvent* (4.0%): payments to *debt management and insolvency companies*.
3. *Insufficient disposable income* (11.6%): average monthly disposable income is less than £100.
4. *Overdraft* (67.4%): one or more days in overdraft (OD) per month for more than 50% of the months throughout their account history.
5. *Returned direct debits* (28.3%): When an applicant has at least one or more returned direct debits (RDD), insufficient funds resulting in a rejection of a pre-arranged payment by the bank.
6. *Gambler* (20.2%): £100 or more on average per month spent on gambling.

Financial behaviours/predictors

Over 100 variables, examples include:

- Financial Management:
  - Inflow (e.g. Total Income, # of income sources, consistency)
  - Outflow (Expenditure by category: housing, groceries, health, gambling)
  - Volatility (**Burstiness in spending**)
- Financial Distress (Debt management payments, overdraft, RDD)
- Financial Resilience (Account Balance, Disposable Income)
- Financial Planning (Insurance, Savings)
- Financial Aid (Total Benefits, Amount by Benefit Types, Pensions)
- Financial Inclusion (# credit providers, # cards, loans, payday loans).

Modelling


Data
Training Set
(80%)
Test Set
(20%)
Out-of-sample
M_{-5}
M_{-4}
M_{-3}
Model training & selection
• Hyperparameter tuning with cross validation (k=10)
• Select optimal model (lowest error)
Period 1
(3 months)
Period 2
(3 months)
Predictive Model
M_{-2}
M_{-1}
M_0
Model testing
• Performance evaluation
Out-of-time

Correlations

		Profile			*Benefits Received*			*Financial Vulnerability Indicators*				
		Age	Employment Length	Female	Carer	Child	Disability	Financial shock	Gambling expenditure	Insolvency & DM	Disposable income	Over-draft
Profile	Employment Length	0.269**										
	Female	0.043	-0.014									
Benefits Received	Carer	-0.001	-0.030	0.038								
	Child	-0.010	0.028	0.362**	0.095**							
	Disability	0.073**	0.011	0.078	0.334**	0.131**						
Financial Vulnerability Indicators	Financial shock withstanding (% of months)	0.182**	0.089**	0.033	0.007	0.053**	0.060**					
	Gambling expenditure (£)	-0.004	0.036	-0.116**	0.008	-0.045**	0.001	0.101				
	Insolvency & debt management expenditure (£)	0.096	0.072	0.008	-0.017	0.022	-0.003	0.103	0.001			
	Disposable income (£)	0.015*	0.027	-0.006	0.043	0.053**	0.051	0.059	0.403**	0.031		
	Overdraft (% of months)	-0.077	-0.014	-0.004	-0.019	0.024	-0.014*	-0.586**	-0.009	-0.026	-0.030	
	No. of returned direct debits	-0.025	-0.043	0.041	0.031	0.088**	0.047	-0.163**	0.025	0.049	0.059	0.332**

Predictive accuracy

		AUROC	
		Mean	*Std*
Withstand financial shock	*LR*	0.756	0.007
	RF	0.903	0.004
	XGB	0.895	0.004
Gambler	*LR*	0.875	0.006
	RF	0.893	0.006
	XGB	0.911	0.006
Returned Direct Debits	*LR*	0.775	0.005
	RF	0.870	0.004
	XGB	0.884	0.004

		AUROC	
		Mean	*Std*
Has a Child	LR	0.824	0.006
	RF	0.917	0.003
	XGB	0.917	0.004
Has a Disability	LR	0.796	0.018
	RF	0.885	0.010
	XGB	0.864	0.010
Female	LR	0.832	0.004
	RF	0.896	0.003
	XGB	0.917	0.002

LR=Logistic regression, RF = Random Forest, XGB = eXtreme Gradient Boosting evaluated with the Area Under the Receiver Operating Characteristic Curve (AUROC), higher values → better predictions.

Impact on protected and sensitive groups

		Acceptance Rate			
		Withstanding financial shock always		No Returned Direct Debits	
		PA Excluded	*PA Included (delta)*	*PA Excluded*	*PA Included (delta)*
Gender	Female	21.4%	0.0%	20.4%	0.0%
	Male	16.1%	0.1%	18.9%	-0.1%
	Male & no benefits	14.4%	-0.3%	18.7%	-0.1%
Benefit	Is carer	25.5%	-0.1%	16.9%	-0.1%
	Has child	23.9%	0.0%	18.7%	0.0%
	Has disability	29.4%	-1.1%	17.7%	-0.1%
Intersectionality	Female & has child	23.5%	0.0%	18.5%	-0.1%
	Female & has disability	31.1%	0.1%	18.1%	0.3%
	Female & is carer	25.7%	-1.4%	16.7%	-0.1%
	Has child & disability	32.9%	-1.5%	16.3%	0.2%
	Has child & is carer	29.4%	0.0%	15.0%	0.2%
	Is carer & has disability	32.0%	0.5%	15.8%	0.2%
Age	Age: 18-21	5.1%	1.2%	16.9%	0.0%
	Age: 22-25	7.2%	0.9%	15.1%	-0.1%
	Age: 26-30	13.1%	1.1%	17.0%	0.3%
	Age: 31-35	18.6%	0.3%	18.1%	0.2%
	Age: 36-40	24.2%	-0.1%	20.1%	0.0%
	Age: 41-45	24.7%	-0.8%	22.2%	0.1%
	Age: 46-50	28.3%	-1.2%	23.3%	0.0%
	Age: 51-55	30.0%	-1.6%	27.2%	-0.8%
	Age: 56-60	32.6%	-1.3%	28.0%	-1.0%
	Age: 60+	30.4%	-0.2%	29.1%	-1.5%

Simulating a scenario, where a lender accepts 20% of applicants, based on the higher probability of e.g. No RDD. Equal outcome would mean 20% acceptance rate across all the subgroups.

Concluding remarks

- We demonstrate the power of OB and AI/ML in profiling the vulnerable customers.
- Good use: e.g. to take early actions to improve the customer's situation.
- Bad use: e.g. targeting vulnerable consumers by predatory lenders.
- We also illustrate how 'proxy' bias arises from hidden associations of seemingly neutral predictors with protected and sensitive attributes.
- There is a need for guidance on dealing with these associations, what level of disparity is acceptable, how to monitor it in absence of protected attributes.
- Intersectionality is another big topic that requires attention and research.
- More details are in the Working Paper.