Machine Learning for Discovery and Prevention:
Identifying behavioural markers of self-reported gambling problems in France and Quebec

W. Spencer Murch, PhD

June 24th, 2022

Disclosure

This research is funded by the Concordia University Chair on Gambling Studies, the Mise sur toi Foundation, Fonds de recherche du Québec – Société et Culture, and The Canadian Institutes of Health Research.

The data were provided by Loto Quebec and the French Online Gambling Regulatory Authority (ARJEL), neither of whom constrain the design, analysis, or publication of our work.

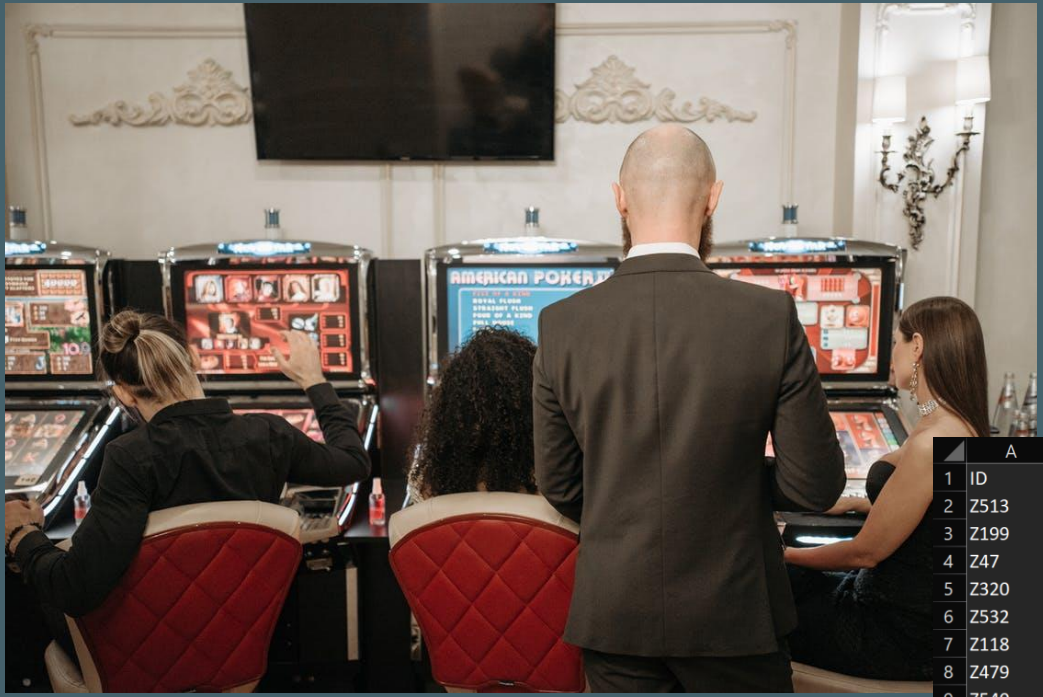
I previously received training and funding from The Centre for Gambling Research at UBC, a research laboratory jointly supported by the Government of British Columbia and the British Columbia Lottery Corporation (BCLC; a Canadian Crown Corporation).

UNIVERSITÉ Concordia UNIVERSITY

CIHR IRSC Canadian Institutes of Health Research Instituts de recherche en santé du Canada

Research chair on gambling studies Chaire de recherche sur l'étude du jeu

Take a moment to imagine…



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	Day	Week	BetPoker	BetSlots	BetLotto	SDPoker	SDSlots	SDLotto	MAXPoker	MAXSlots	MAXLotto	Logins	Cashouts
2	Z513	5	47	90	116	0	13.06	402.99	0.31	69	93	2	3	1
3	Z199	3	3	73	26	0	14.97	72.59	0.44	38	82	2	2	2
4	Z47	3	28	5	68	1	19.89	339.37	0.45	56	147	2	3	0
5	Z320	3	13	95	129	0	15	2.37	0.85	139	41	2	2	1
6	Z532	6	3	1	151	0	1.3	176.73	0.68	23	150	1	0	1
7	Z118	1	33	9	181	1	12.7	21.43	0.34	131	56	1	3	1
8	Z479	6	36	81	145	0	19.65	270.55	0.33	128	58	2	1	1
9	Z540	4	30	55	34	0	16.95	171.26	0.65	135	168	0	2	1
10	Z682	7	14	90	94	0	2.06	3.94	0.8	37	120	3	1	1
11	Z19	3	31	11	101	0	5.07	144.15	0.06	26	78	2	2	0
12	Z139	5	32	2	8	1	2.25	182.22	0.35	53	166	1	1	1
13	Z344	3	1	5	76	1	23.86	46.33	0.18	75	150	3	0	1
14	Z23	1	17	85	136	0	17.86	116.5	0.06	50	144	0	1	2
15	Z304	2	2	16	86	0	7.06	481.25	0.38	38	143	2	1	2
16	Z619	5	44	27	87	1	24.58	364.44	0.65	15	80	1	2	2
17	Z534	7	36	22	77	1	6.94	517.3	0.5	38	44	3	2	2
18	Z561	4	7	89	136	0	14.33	35.86	0.01	34	115	2	1	0

Gambling is changing

LA PRESSE+

May 16th, 2022

LOTO-QUEBEC

THE NUMBER OF SELF-EXCLUDED EXPLODES

More and more players are demanding that their access to online gambling sites be closed

VINCENT LARIN
THE PRESS

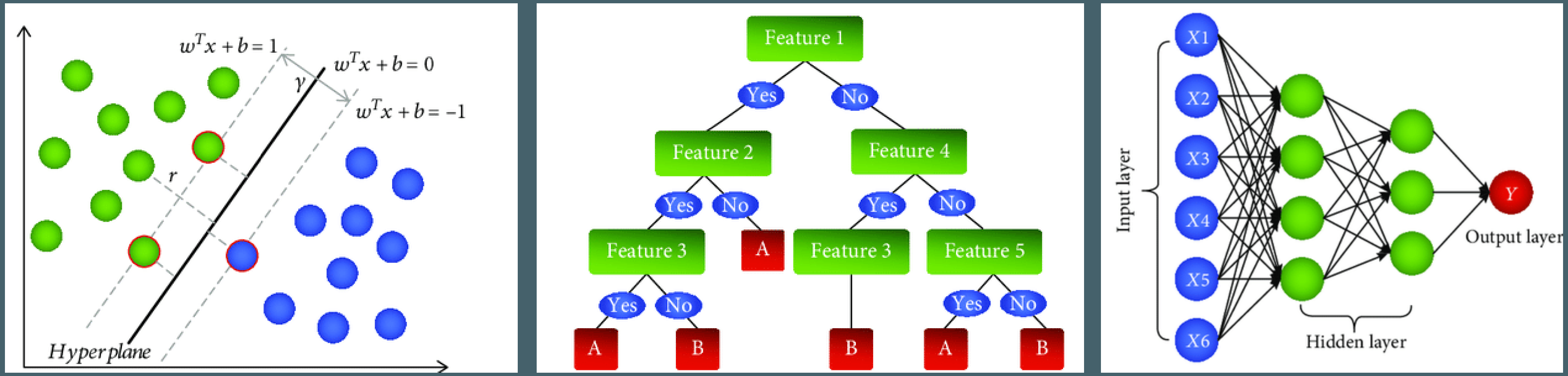
The number of registrations for Loto-Québec's self-exclusion program for online gambling has increased markedly during the pandemic. *(72% from 2020 to 2021)*

Gambling harm prevention needs detection systems

- How can you detect people at-risk for experiencing harm?
  - Ask all users to fill out a questionnaire every week? **TOO FATIGUING.**
  - Have someone look through user accounts on a case-by-case basis? **TOO NOISY.**
  - Intervene after a particular threshold (e.g. $500 in losses) is crossed? **NOT PERSONALLY-RELEVANT.**

Machine learning

- Machine learning provides a suite of statistical tools specifically designed to make good predictions about future events



(Figure from Yin et al., 2020)

Machine learning for gambling harms

- Machine learning approaches have shown good performance in classifying relevant indicators of harm:
  - Triggering RG alerts (Gray et al., 2012)
  - Self-exclusion from gambling (Finkenwirth et al., 2020; Haefeli et al., 2015; Percy et al., 2016)
  - Account closure due to self-reported problems (Braverman & Shaffer, 2012; Philander, 2014; Xuan & Shaffer, 2009)
  - Brief Biosocial Gambling Screen (partial DSM criteria) (LaPlante et al., 2014)
  - Problem Gambling Severity Index (PGSI) (Luquiens et al., 2016)

Research aims

1. Determine if (and *which*) machine learning algorithms can accurately predict PGSI risk levels using data that is readily available to online gambling websites

2. Explore potential behavioural markers of gambling harm

Study 1 - France

- N = 9,306 users of licensed, privatized gambling websites

- PGSI email invitations sent to users from:

  December 2015 – March 2016

- Gambling data provided by *French Online Gambling Regulatory Authority* (*ARJEL*) for 12 months prior to each user’s PGSI

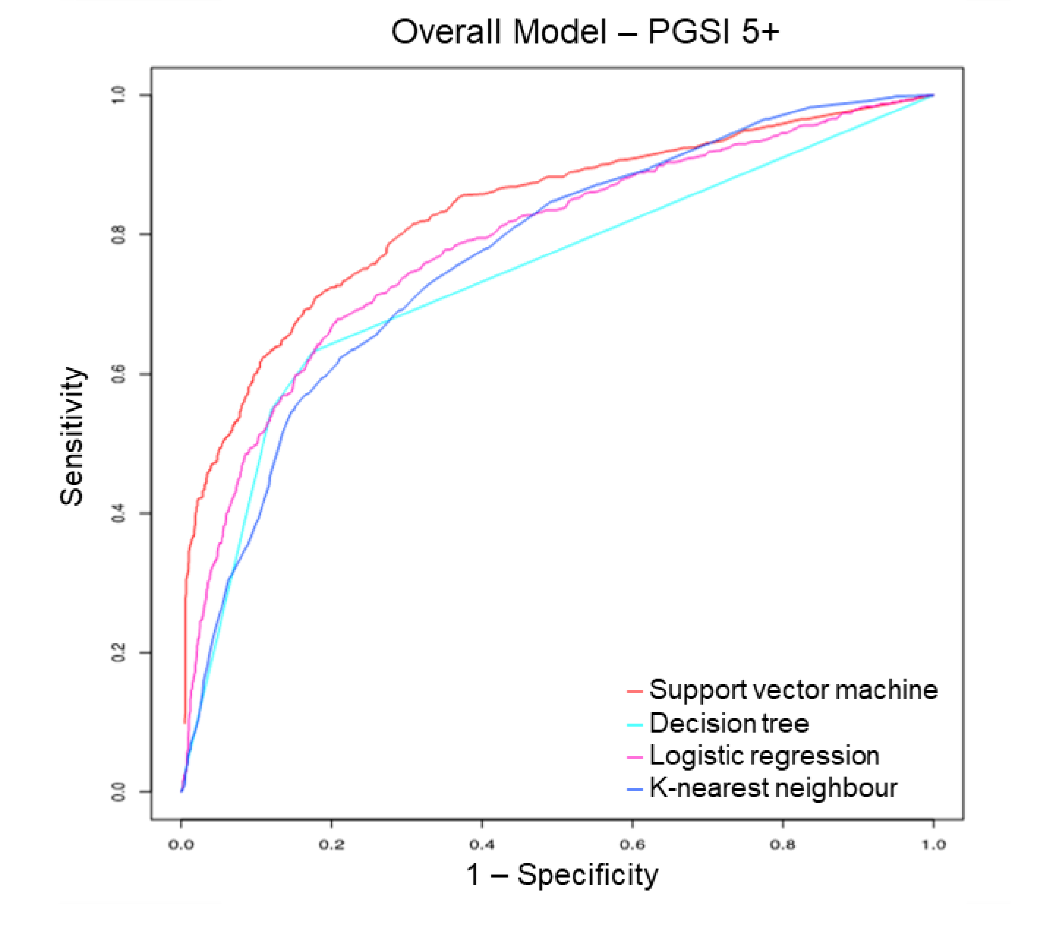
Study 1 - Methods

- **Input Variables (64)**
    1. Account-level information (e.g., age, sex)
    1. Usage of RG tools
    2. Deposits and withdrawals
    3. Betting Information
    4. Loss chasing

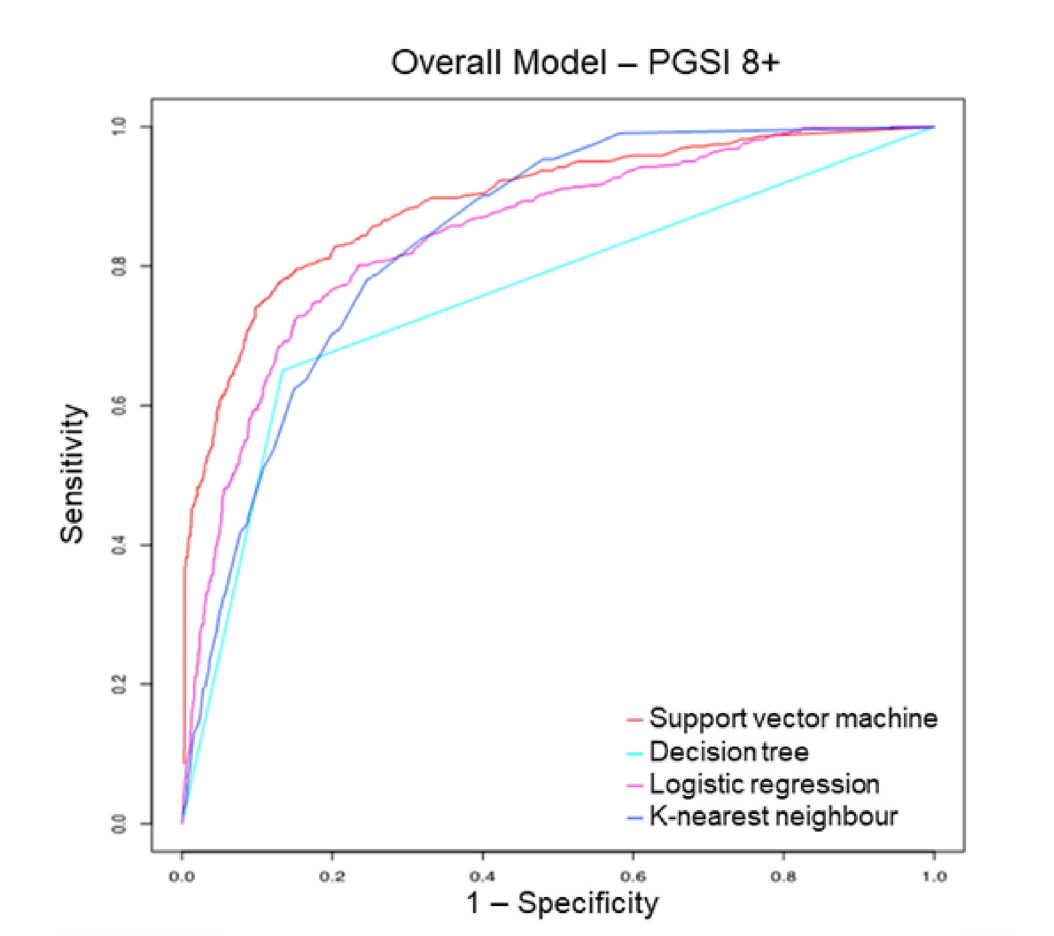
- **Dependent Variables (2)**
    1. PGSI 5+: Moderate-to-high risk of experiencing past-year gambling harms
    2. PGSI 8+: High risk of experiencing past year gambling harms

- **ML Algorithms (4)**
    1. Logistic regression
    2. K-nearest neighbours
    3. Decision Trees
    4. Support Vector Machines

Study 1 – ROC plots



AUC = 83.20%



AUC = 87.70%

Study 1 – Performance

Metric	PGSI 5+	PGSI 8+
Sensitivity	71.00%	74.30%
Specificity	82.10%	87.20%
Positive Predictive Value (PPV)	49.62%	38.67%
Negative Predictive Value (NPV)	91.89%	96.87%

Study 2 - Quebec

- N = 9,145 users of *lotoquebec.com* (formerly *espacejeux.com*)

- Email invitations sent to users from:

  September 2019 – November 2019

- Gambling data provided by *Loto Quebec* for 12 months prior to each user's PGSI

Study 2 - Methods

♦ **Input Variables (144)**

1. Account-level information (e.g., age, sex)
1. Usage of RG tools
2. Deposits and withdrawals
3. Betting Information
4. Loss chasing

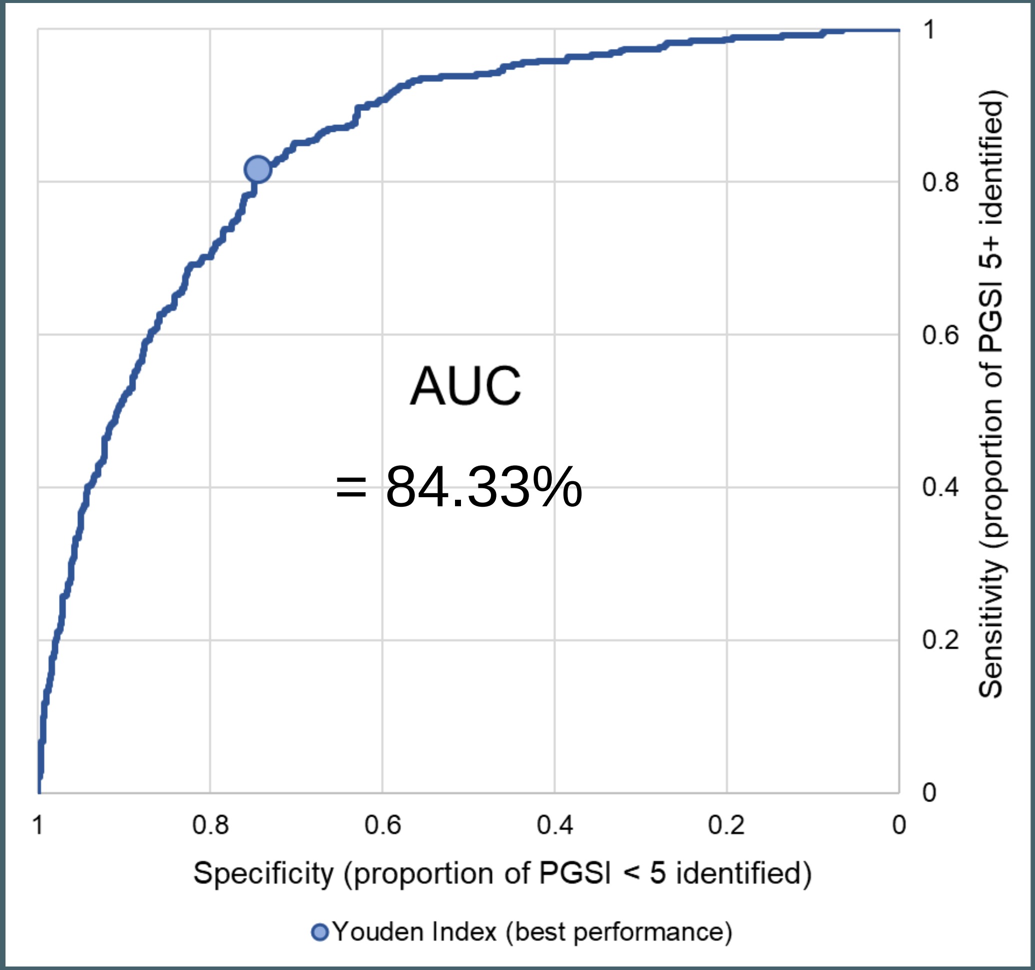
♦ **Dependent Variables (2)**

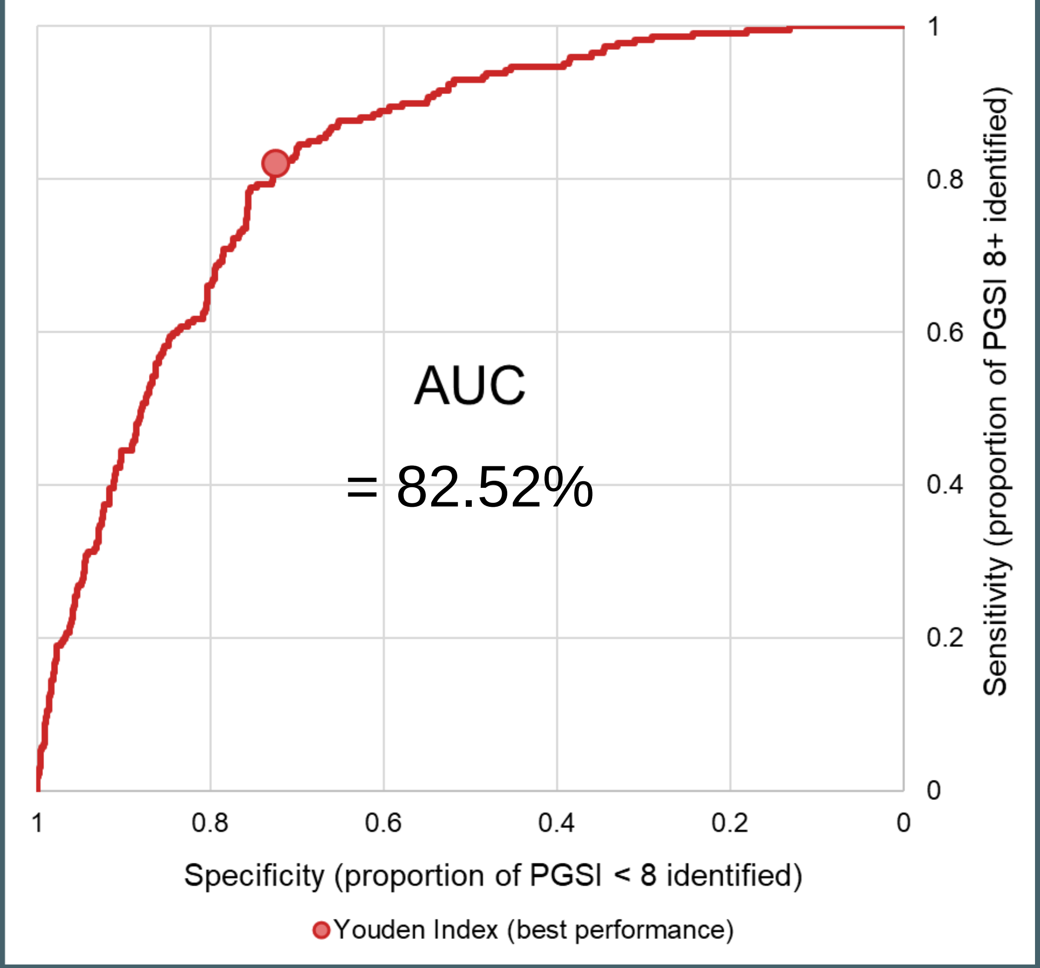
1. PGSI 5+: Moderate-to-high risk of experiencing past-year gambling harms
2. PGSI 8+: High risk of experiencing past year gambling harms

♦ **ML Algorithms (6)**

1. Logistic regression
2. K-nearest neighbours
3. Decision Trees
4. Support Vector Machines
5. Random Forest *Added*
6. Artificial Neural Networks *Added*

Study 2 – ROC plots

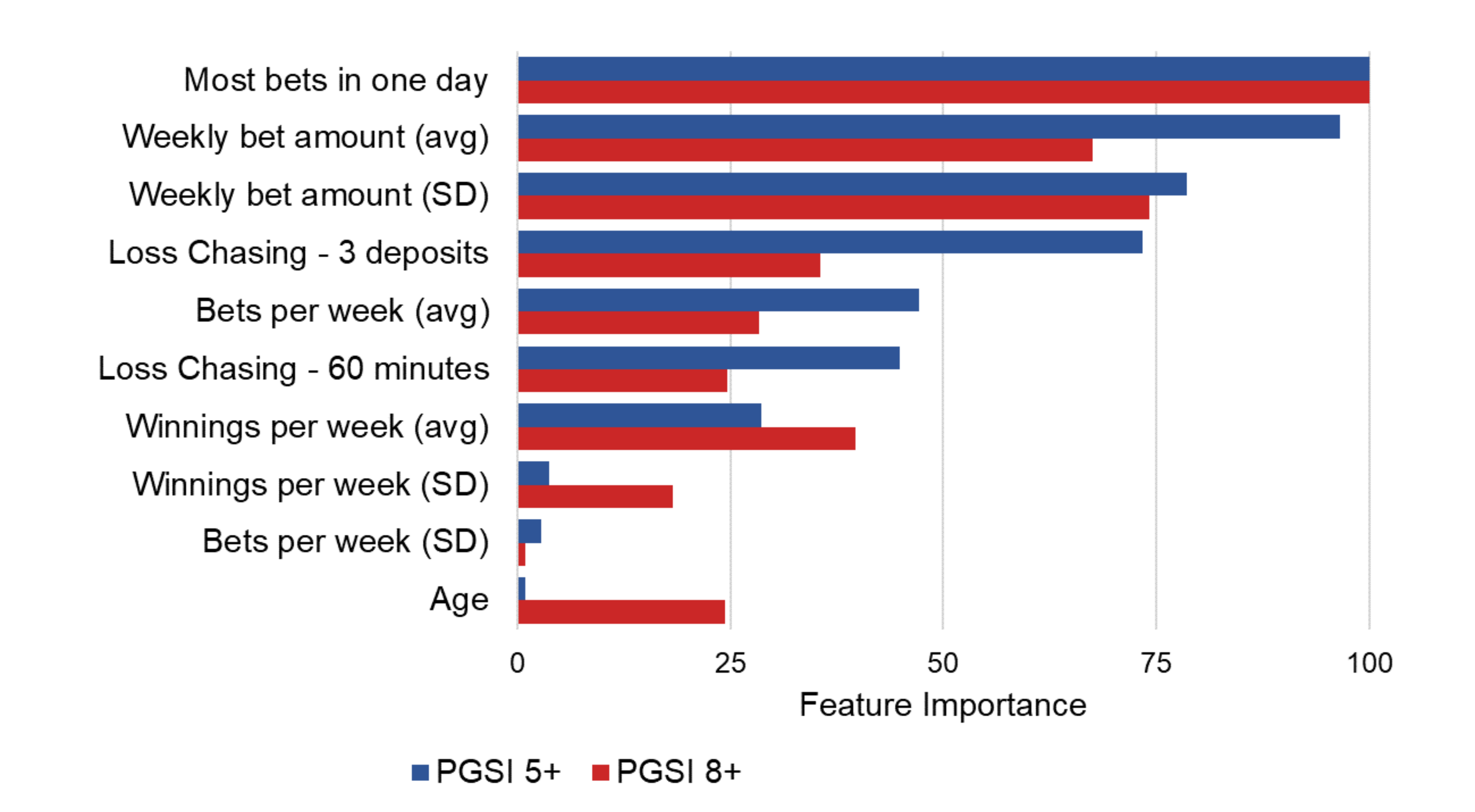




Study 2 – Performance

Metric	PGSI 5+	PGSI 8+
Sensitivity	81.75%	81.94%
Specificity	74.36%	72.20%
Positive Predictive Value (PPV)	46.29%	29.48%
Negative Predictive Value (NPV)	93.78%	96.57%

Study 2 – relative importance



General conclusions

- Machine learning algorithms can provide excellent classification performance for PGSI risk categories using online gambling data
- Specific aspects of betting behaviour distinguish users at different risk levels
- Aggregate risk-detection for identifying more-harmful situations

Next steps:

1. Re-validate models (N = 13,300)
2. Achieve equitable outcomes
3. Deployment online
4. Evaluate interventions *in the place they would be used*

Thank you!

Dr. Sylvia Kairouz

Dr. Martin French

Coauthors

Dr. Jean-Michel Costes,
Sophie Dauphinais,
Elyse Picard,
Vincent Eroukmanoff,
Clément Carrier,
Dr. Pascal Doray-Demers

UNIVERSITÉ Concordia UNIVERSITY

CIHR IRSC Canadian Institutes of Health Research Instituts de recherche en santé du Canada

Research chair on gambling studies
Chaire de recherche sur l'étude du jeu

Disclosure

This research is funded by the Concordia University Chair on Gambling Studies, the Mise sur toi Foundation, Fonds de recherche du Québec – Société et Culture, and The Canadian Institutes of Health Research.

The data were provided by Loto Quebec and the French Online Gambling Regulatory Authority (ARJEL), neither of whom constrain the design, analysis, or publication of our work.

I previously received training and funding from The Centre for Gambling Research at UBC, a research laboratory jointly supported by the Government of British Columbia and the British Columbia Lottery Corporation (BCLC; a Canadian Crown Corporation).

Université Concordia University

CIHR IRSC Canadian Institutes of Health Research Instituts de recherche en santé du Canada

Research chair on gambling studies Chaire de recherche sur l'étude du jeu

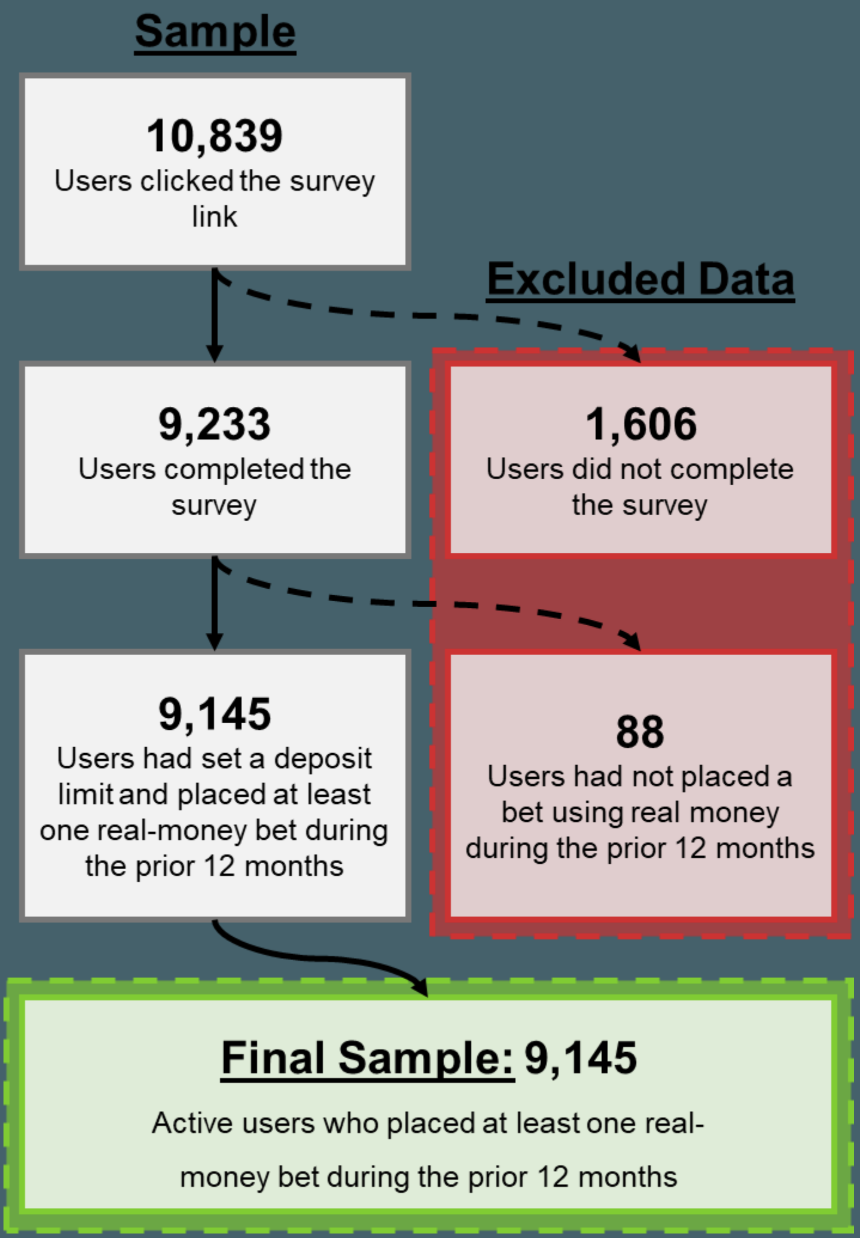
Study 1 - results

Higher-risk gamblers correctly classified	Percentage of lower-risk gamblers correctly classified	
	PGSI 5+	PGSI 8+
99%	4.90%	11.90%
95%	24.70%	40.10%
90%	44.20%	58.80%
85%	63.40%	71.70%

Study 2 - results

Higher-risk participants correctly classified	Percentage of correctly-classified lower-risk participants	
	PGSI 5+	**PGSI 8+**
99%	13.62%	24.36%
95%	45.93%	39.16%
90%	61.78%	55.03%
85%	70.26%	68.64%

Study 2 - results



Study 2 - results

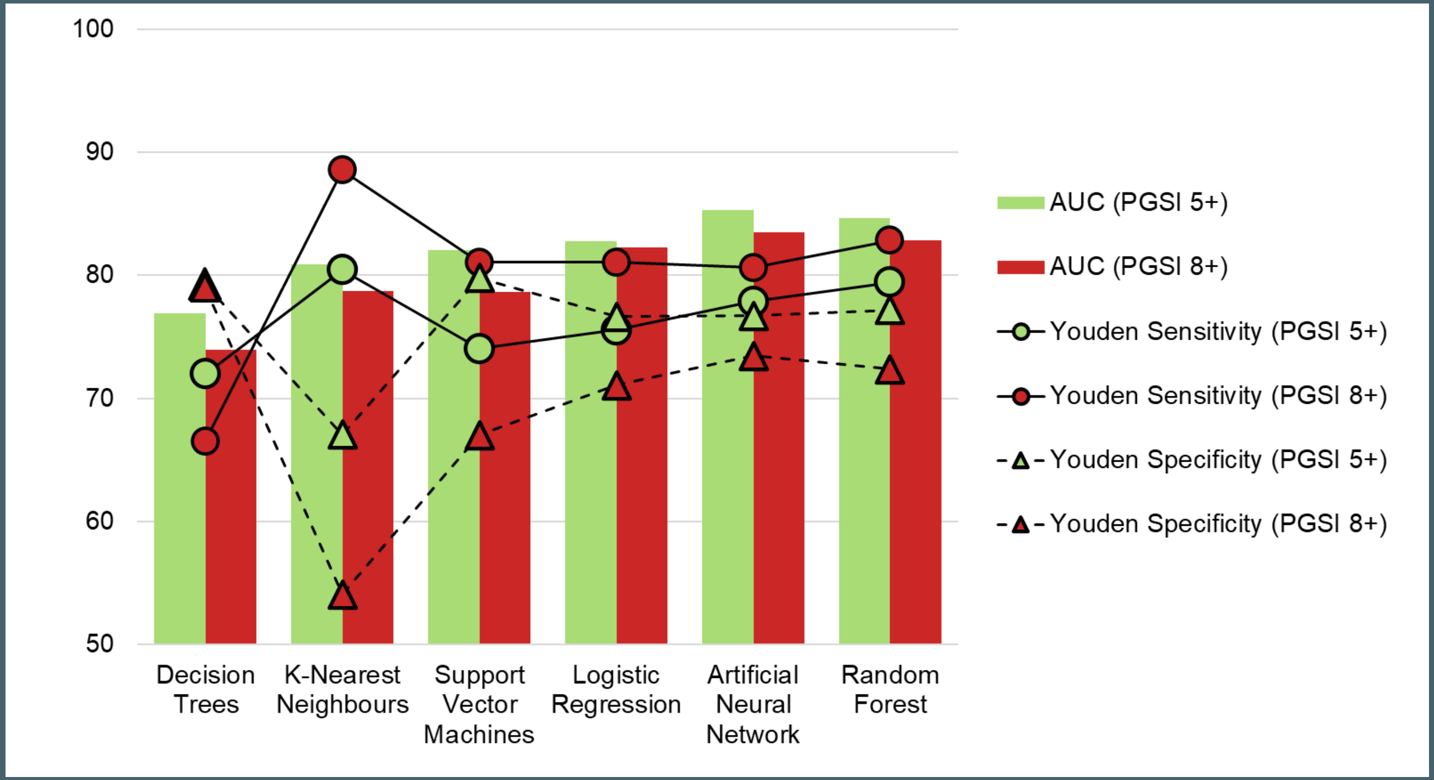
Models

Variable Group	Number of Variables in the Group	Model 1	Model 2	Model 3a	Model 3b	Model 4	Model 5	Model 3a Reduced
Account information	5	■	■	■	■	■	■	■
Responsible gambling tool use	8		■	■	■	■	■	
Loss chasing	2			■	■	■	■	■
Overall gambling activity	10			■		■	■	■
Transactions	19				■	■	■	
Individual gambling activities	100						■	
Number of variables in the model		5	13	25	34	44	144	10
Model performance (AUC)		Good	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent

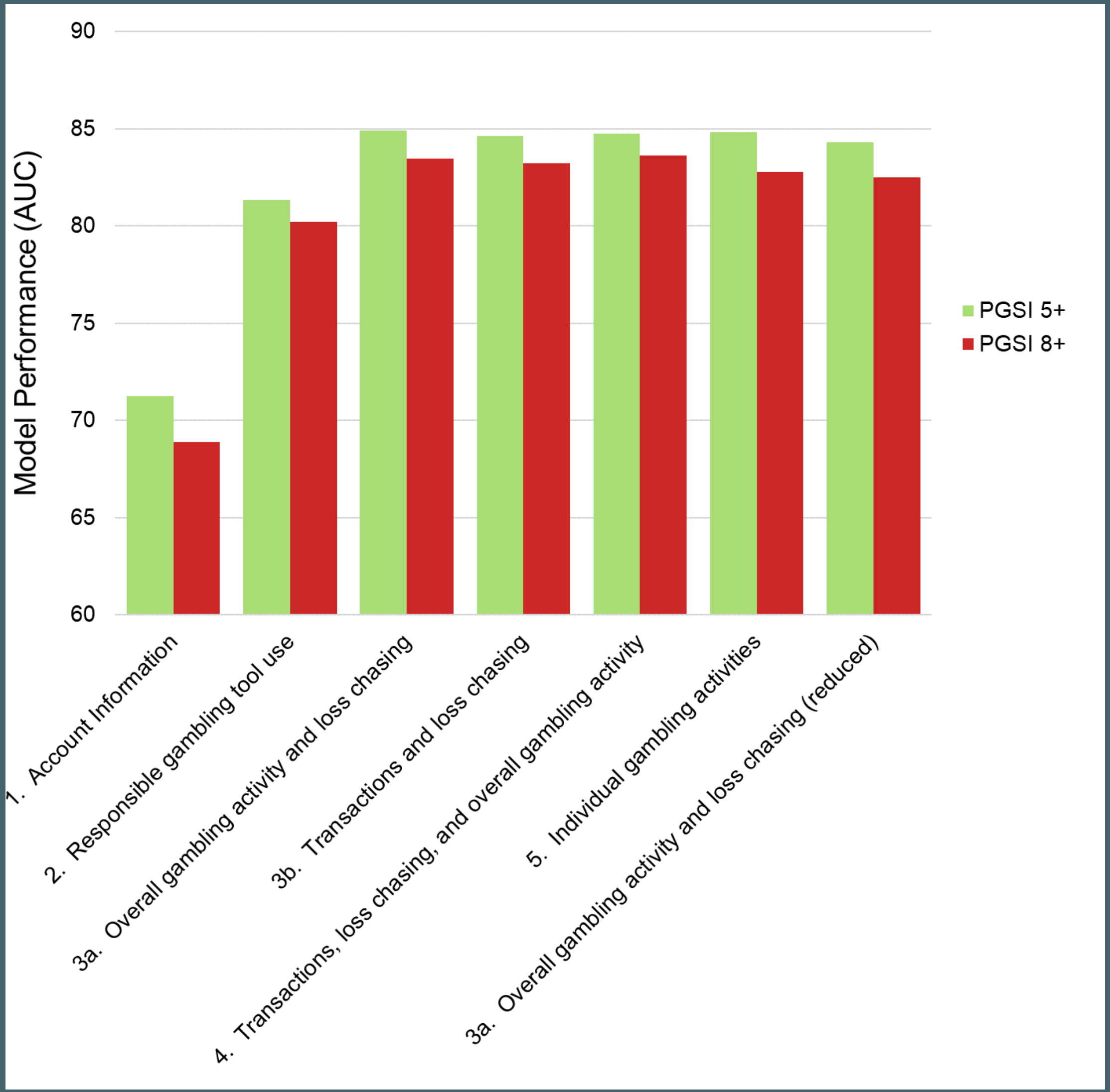
Performance

Poor — Excellent

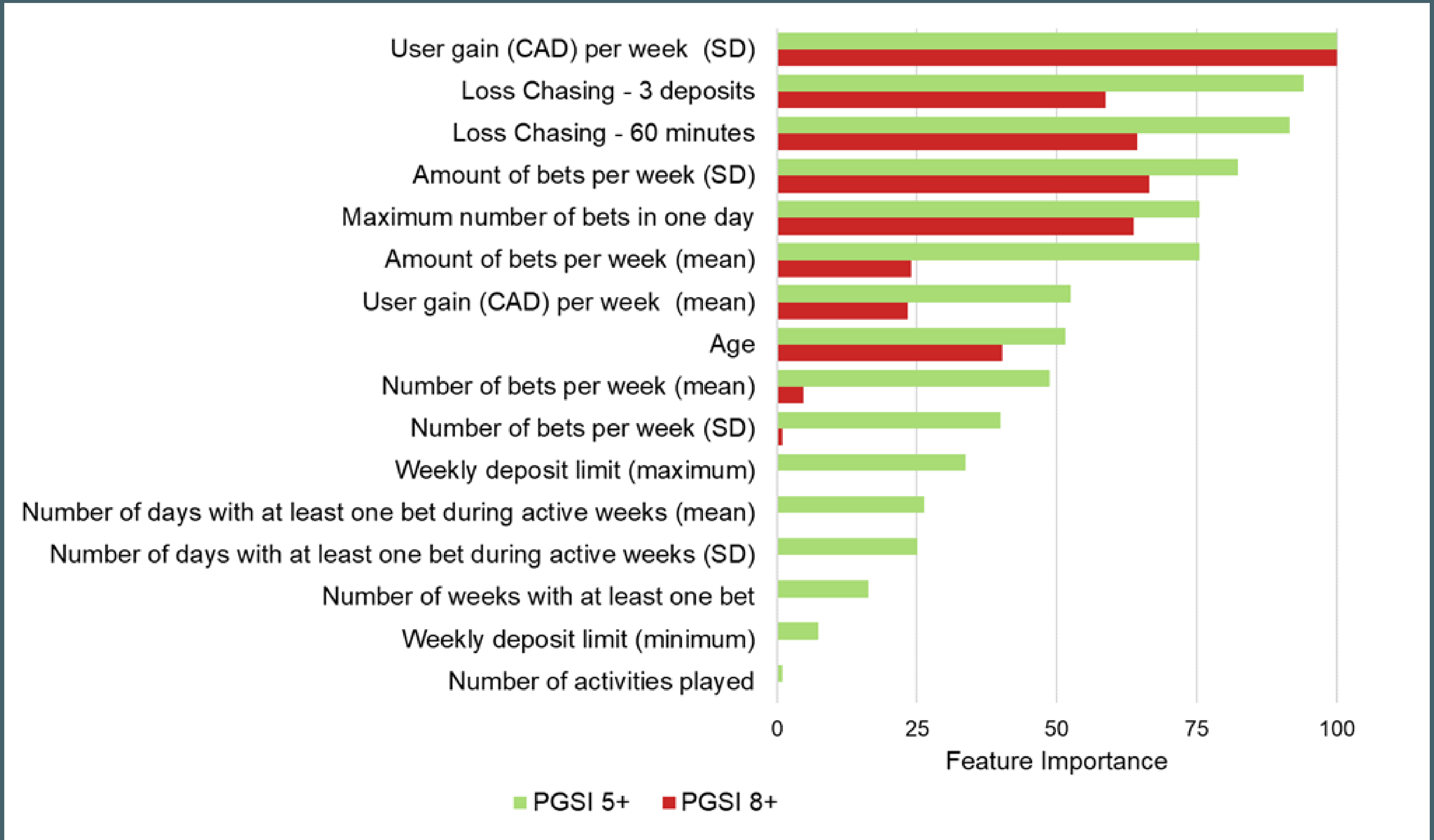
Study 2 - results



Study 2 - results



Study 2 – results (slot machines)



Study 2 – results (lottery)