1828
AKADÉMIAI KIADÓ

Predicting problem gambling among online sports and race bettors: Assessing the value of machine learning using behavioural and self-reported data

Journal of Behavioral Addictions

15 (2026) 2, 846–862

DOI:
10.1556/2006.2025.00525

ROBERT M. HEIRENE[1*], EDEN ZHANG[2], DARYA VANICHKINA[2], CHARLES T. DE LEAU[1,3], EUNICE L. Y. HUYNH[1] and SALLY M. GAINSBURY[1]

[1] Brain & Mind Centre, School of Psychology, University of Sydney, NSW, Australia

[2] Sydney Informatics Hub, University of Sydney, NSW, Australia

[3] Development and Psychopathology (ADAPT)-Lab, Department of Psychology, University of Amsterdam, The Netherlands

Received: November 10, 2025 • Revised manuscript received: May 6, 2026 • Accepted: May 7, 2026
Published online: June 2, 2026

FULL-LENGTH REPORT

Check for updates

ABSTRACT

Background and aims: Online gambling operators collect detailed behavioural data that can identify customers at risk of harmful gambling. However, there is limited clarity on how to optimally achieve this in practice, including which variables are most useful and whether short-term data windows are sufficient for risk detection. These details are increasingly important as regulatory frameworks emphasise timely intervention. We examined the value of machine learning in this context by comparing models trained on 30 days versus six months of behavioural data and exploring whether incorporating survey responses enhanced performance. *Methods:* Customers from two Australian sports and race betting sites (N = 1,470) completed a survey including the Problem Gambling Severity Index (PGSI) and measures of employment, income, gambling satisfaction, and number of gambling accounts. We built machine learning models to classify participants into risk groups (PGSI 1–7 [no-to-moderate-risk] vs. PGSI≥8 [high-risk]), comparing performance across data windows (30 days vs. six months), and with or without survey variables. *Results:* Models using only behavioural data achieved adequate classification accuracy (AUROC = 0.74–0.75), with similar performance across 30-day and six-month windows. The most predictive account-based variables were age, deposits per active day, average stake, and days since betting. Combining behavioural data with self-reported variables enhanced performance (AUROC = 0.76–0.85). Two self-reported variables—number of gambling accounts held and gambling satisfaction—were primarily responsible for these improvements. *Conclusions:* Machine learning models can detect at-risk customers on online sports and race betting sites using only 30 days of behavioural data. Performance can be improved by adding minimal, non-intrusive self-report measures.

KEYWORDS

gambling disorder, risk detection, artificial intelligence, algorithm, behavioural marker

INTRODUCTION

Online gambling accounts for a large and growing proportion of the global gambling market and, in some jurisdictions, has surpassed land-based gambling (European Gaming & Betting Association, 2025). In Australia, more than one in ten people report gambling online (ACMA, 2022). Participation in online gambling, compared to traditional land-based forms, has been linked to higher rates of problematic gambling (Allami et al., 2021); although the highest rates are found among those who combine both forms (Hing, Russell, et al., 2022).

*Corresponding author.
E-mail: robert.heirene@sydney.edu.au

AKJournals

Online gambling harms are attributed to structural features relating to the speed, accessibility, ease of betting; the lower salience of digital spending; aggressive marketing and frequent promotions; and, in many jurisdictions, less regulation and consumer protections than land-based gambling (Hing et al., 2024; Hing, Smith, et al., 2022).

Unlike anonymous cash-based gambling, online gambling operators collect detailed data on all customer transactions and site interactions (Delfabbro, Parke, & Catania, 2023; Gainsbury, 2011). This behavioural account or tracking data offers a powerful tool for detecting customers at risk of problematic or harmful gambling and could enable timely, personalised interventions (Auer & Griffiths, 2014; Gainsbury, 2011; LaBrie & Shaffer, 2011). Artificial Intelligence (AI), including machine learning algorithms, is being increasingly used in this risk detection context (Ghaharian et al., 2023; Marionneau, Ristolainen, & Roukka, 2025). Machine learning presents an efficient and economical way to process large datasets of behavioural account data and identify patterns across multiple behavioural markers of harm (e.g., time and money spent). Machine learning models can distinguish between lower- and higher-risk users, supporting operators in delivering targeted harm-reduction strategies.

Licensing bodies typically require online operators to provide a voluntary self-exclusion (VSE) feature that allows customers to request that operators block them from gambling with their site(s) (Gainsbury, 2014). Studies applying machine learning for risk detection often use VSE as a proxy for problem or high-risk gambling as their target or outcome variable. This assumes that VSE is primarily motivated by gambling-related harms (Catania & Griffiths, 2021). Studies using VSE as their outcome have obtained adequate to excellent classification performance[1] and often show that Random Forest (a commonly used machine learning algorithm that combines multiple decision trees to make predictions) outperforms other approaches (Finkenwirth, MacDonald, Deng, Lesch, & Clark, 2020; Hopfgartner, Auer, Griffiths, & Helic, 2022; Percy, França, Dragičević, & d'Avila Garcez, 2016). VSE captures a clearly defined group and is easily recorded by operators; however, it is an unreliable proxy for problem gambling (Griffiths & Auer, 2016). Individuals may self-exclude for reasons that do not include the experience of gambling harms, including to prevent harms or to simply stop using the site (Hayer & Meyer, 2011). Catania and Griffiths (2021) found half of the customers who used VSE did so within the first seven days of registration and spent less money than those who closed their accounts due to gambling problems. Thus, models built to predict VSE may subsequently fail to identify individuals who are genuinely at risk of experiencing gambling problems.

Recent machine learning studies have used self-reported problem gambling scores to provide a more reliable indication of risk than VSE. Combining self-report customer survey and behavioural account datasets is increasingly feasible with partnerships between researchers and operators, in some cases facilitated by gambling regulators (Delfabbro et al., 2023; Forrest & McHale, 2022). To date, at least six studies have built models using account data to detect self-reported problem gambling status as determined via the Problem Gambling Severity Index (PGSI; Ferris & Wynne, 2001). These studies (Auer & Griffiths, 2022; Hopfgartner, Auer, Helic, & Griffiths, 2024; Kairouz et al., 2023; Luquiens et al., 2016; Murch et al., 2023; Perrot et al., 2022) have obtained adequate–good classification performance (Area Under the Receiver Operating Characteristic Curve [AUROC] = 0.72–0.88) and similarly find that Random Forest models perform well.

While the understanding of online gambling risk detection has been advanced by studies linking behavioural and survey datasets, further efforts are needed to understand how machine learning models can be optimised in this context. In a recent priority setting exercise, using "big data" and approaches such as machine learning to better understand indicators of risky/harmful gambling and how we can identify and support at-risk consumers was raised as an important research priority and ultimately ranked 22nd (out of 41) in important topics relating to interventions. Others have also recently highlighted the need for more research that can assist with the early identification of at-risk individuals (Bowden-Jones et al., 2022). Several jurisdictions now require operators to actively monitor their customers for signs of problematic gambling, including the UK (UK Gambling Commission, 2022) and Sweden (Lakew & Lindner, 2025). This demonstrates the need for research that can guide the implementation of effective risk detection systems (for a recent overview of countries requiring customer risk monitoring, see Auer & Griffiths, 2026).

The variation in the performance of machine learning models by the data windows available to train the models (e.g., one week vs. one year of behavioural data) remains an area in need of investigation. One study of 35,048 Swedish online casino gambling customers demonstrated that models trained on different behavioural data timeframes (30, 60, and 90 days) performed similarly, with marginally better performance for the shorter timeframes (Andersson, Carlbring, Lyon, Bermell, & Lindner, 2025). This is important as identifying risky gambling early in a person's account history can prompt earlier interventions. However, participants in the study were given non-precise labels (higher- or lower-risk) based on assessments conducted by the gambling operator. Verification of these classifications was not included in the study and they appear to have been partially based on the same variables used in the classification model (e.g., deposit patterns, session duration). This creates a risk of circularity, wherein the accuracy of model predictions is inflated as they were informed by the same information used to establish initial classifications. Further studies have not compared different windows of behavioural data for model

[1]As defined by Area Under the Receiver Operating Characteristic Curve (AUROC) values in the range of 0.75–0.94. See Statistical analysis section for a description of this metric.

development, typically focusing on a 12-month period (Murch et al., 2023; Perrot et al., 2022). In practice, waiting to obtain protracted behavioural data before applying these models may delay the identification of at-risk customers, thus contributing to greater harms. Contemporary regulatory frameworks now require many operators to identify and respond to risk over short timeframes. For example, the UK Gambling Commission requires operators to continuously monitor customer behaviour and act in a timely manner, which may include real-time interventions when indicators of harm are observed (UK Gambling Commission, 2023).

Self-reported data from consumers provides valuable context related to risk of problematic gambling unavailable in operator-held data (Forrest & McHale, 2022). The integration of self-reported data into machine learning models has not yet been well researched. Recent studies indicate that this approach may hold value. Auer and Griffiths (2023) developed a linear model to predict responses to a single item ("*My gambling is a problem sometimes*") by combining a small number of self-reported and account variables from race betting consumers in Norway (N = 3,627). Sacco and Jeong (2025) combined ticket upload data and survey responses from 5,903 US lottery players, finding that the six most important variables in their model were all self-reported, including frequency of other gambling, income, education, age, and employment status. Overall, this nascent approach requires more extensive investigation.

To our knowledge, no studies have developed machine learning models for identifying at-risk consumers on Australian sites or specifically in an online sports and race betting context. Evidence suggests that regional differences in gambling behaviour may impact model efficacy (Hopfgartner et al., 2024). Australia represents a unique jurisdiction with high per capita gambling expenditure (Greer et al., 2023) and an online market restricted to sports and race betting and lottery. Existing machine learning studies attempting to predict self-reported problem gambling have focused on online casino (Auer & Griffiths, 2022; Hopfgartner et al., 2024), poker (Luquiens et al., 2016), or mixed modalities (Murch et al., 2023; Perrot et al., 2022), precluding specific insights into the unique risk behaviours in a sports and race betting environment (e.g., variability in sports bet on, odds selections). The performance of these models in an Australian sports and race betting context, and which variables are most predictive of risky play, remains unclear. In this study, we aimed to extend the existing literature on online gambling risk detection by:

1. building machine learning models to predict self-reported problem gambling using behavioural account data from Australian sports and race betting consumers
2. comparing the classification performance of models built using past-30-day versus past-six-month behavioural account data
3. comparing the classification performance of models built with and without survey-based variables.

To enhance the practical relevance of results, survey-based variables were included only if they satisfied two criteria: [1] the information could be plausibly collected by operators or regulators with minimal disruption to consumers, which was operationalised as being obtainable via a single question, and [2] the information must have an obvious connection with risk, such that consumers could easily understand the request without needing additional explanation (e.g., income, but not marital status).

METHODS

No hypotheses or data analysis plans were preregistered as this was an exploratory study. Whilst we planned to use the data for this purpose before collection, the methods were finalised after observing the data. The overall methods used were, however, preregistered on Open Science Framework (https://osf.io/vdsmw), including the recruitment strategy, participant eligibility, data collection survey, and variables collected.

Participants and procedures

Customers from two online sports and race wagering sites in Australia owned by the same parent company (Entain Australia) were invited to take part in a survey about their gambling. Eligible customers had to have placed at least one bet with the site in the preceding six months, held an account for at least 30 days, and could not have a suspended account, be self-excluded, or be on a timeout/take-a-break. Two cohorts from the total eligible population were identified for inclusion in the study by the operator. These included all customers flagged by the operator's internal risk-detection system as "at-risk" in the six months prior and a random sample of all "not-at-risk" customers at a 20:80 *at-risk/not-at-risk* ratio. This resulted in a sample of 4,829 "at-risk" and 20,000 "not-at-risk" customers (rounded to the nearest thousand). These 24,829 customers were invited to take part in a survey about their gambling in January 2024. Email invitations were sent via the operator, and an SMS reminder was sent eight days later. In total, 3,867 people opened the survey and 1,470 answered all questions. Participants were given the option to enter a prize draw to win one of 20 e-gift vouchers valued at $250, which could not be redeemed for cash or spent on gambling.

In a separate analysis of the data used here, we showed that the survey sample (i.e., the 1,959 people completing all PGSI items) was broadly representative of the wider population invited to take part, but was more engaged on average (Heirene, Cobb-Clark, Tymula, Santos, & Gainsbury, 2025). Most notably, survey responders bet more regularly in the preceding six months than non-responders (Cohen's $d = 0.58$), had placed a bet more recently ($d = -0.51$), were older ($d = 0.40$), and made more deposits ($d = 0.32$). A summary of the demographic characteristics of the two samples used in our analyses are presented in Table 1.

Measures & feature generation

Outcome variable. The Problem Gambling Severity Index (PGSI; Ferris & Wynne, 2001) was used to assess

Table 1. Sample demographic characteristics

Characteristic	Full sample $N = 1{,}470$	Subsample with 6 months of betting $N = 1{,}349$
Age	40 (30, 52)	40 (30, 53)
Gender		
Female	154 (10%)	135 (10%)
Male	1,292 (88%)	1,190 (88%)
Unknown*	24 (1.6%)	24 (1.8%)
Employment		
Employed	1,384 (94%)	1,273 (94%)
Unemployed	86 (5.9%)	76 (5.6%)
Education#		
Year 11 or below (includes Certificate I/II/n)	246 (17%)	224 (17%)
Year 12	359 (24%)	329 (24%)
Certificate III/IV	255 (17%)	236 (17%)
Advanced diploma/diploma	128 (8.7%)	116 (8.6%)
Bachelor's degree	276 (19%)	256 (19%)
Graduate diploma or graduate level certificate	117 (8.0%)	105 (7.8%)
Master's degree	76 (5.2%)	72 (5.3%)
Doctoral degree	13 (0.9%)	11 (0.8%)
Household income (pre-tax)		
Negative or Zero Income	11 (0.7%)	8 (0.6%)
$1–$49,999/yr ($1–$959/wk)	169 (11%)	151 (11%)
$50,000–$59,999/yr ($960–$1,149/wk)	72 (4.9%)	66 (4.9%)
$60,000–$79,999/yr ($1,150–$1,529/wk)	148 (10%)	136 (10%)
$80,000–$99,999/yr ($1,530–$1919/wk)	177 (12%)	163 (12%)
$100,000–$124,999/yr ($1920–$2,399/wk)	166 (11%)	150 (11%)
$125,000–$149,999/yr ($2,400–$2,879/wk)	120 (8.2%)	116 (8.6%)
$150,000–$199,999/yr ($2,880–$3,839/wk)	194 (13%)	184 (14%)
$200,000+/yr ($3,840+/wk)	231 (16%)	210 (16%)
Don't know	28 (1.9%)	27 (2.0%)
Prefer not to say	154 (10%)	138 (10%)
PGSI score	3 (1, 7)	3 (1, 7)
Number of gambling accounts		
1	264 (18%)	238 (18%)
2	441 (30%)	398 (30%)
3	363 (25%)	337 (25%)
4	179 (12%)	169 (13%)
5+	223 (15%)	207 (15%)
Days since site registration	1,211 (510, 2,182)	1,333 (694, 2,254)
Days since last bet on site	4 (2, 28)	4 (2, 27)

Values presented: continuous variables: Median (IQR); categorical variables: N (percent of sample).
*Gender not reported to operator.
#Ordered from lower to higher levels of education.

self-reported problem gambling via the survey. The PGSI consists of nine items assessing problem gambling behaviours and adverse consequences of gambling (Ferris & Wynne, 2001). Respondents rate their agreement with the nine statements on a scale from 0 (Never) to 4 (Always). It is possible that some participants may have provided inaccurate or rushed responses to the PGSI and other survey questions. Whilst we did include an attention check early in the survey in the Gambling Harm Measure (GHM) (participants were asked to select "Yes" to the question, which for the other GHM items would indicate experiencing that harm), we chose not to exclude participants who did not pass this for two reasons. First, 78% of those who failed the GHM attention check scored 0 on all items of the GHM, suggesting that they defaulted to responding "No" to all items to indicate the absence of harm. Second, we ran our first model (see Fig. 2) with and without those who failed the attention check included, finding negligible differences in model performance (AUROC = 0.748 with them excluded, AUROC = 0.752 with them included) and variable importance (see Supplemental Figure S1 for the full outcomes from this model). Histograms of the PGSI scores for the two samples used in our analyses are presented in Fig. 1, showing no clear abnormalities in distribution (e.g., large numbers of people scoring the maximum value).

Predictor variables. Behavioural features from account data: The operator provided account data for all 24,879 customers invited to the survey for a period spanning six months before and after the survey. Account data included

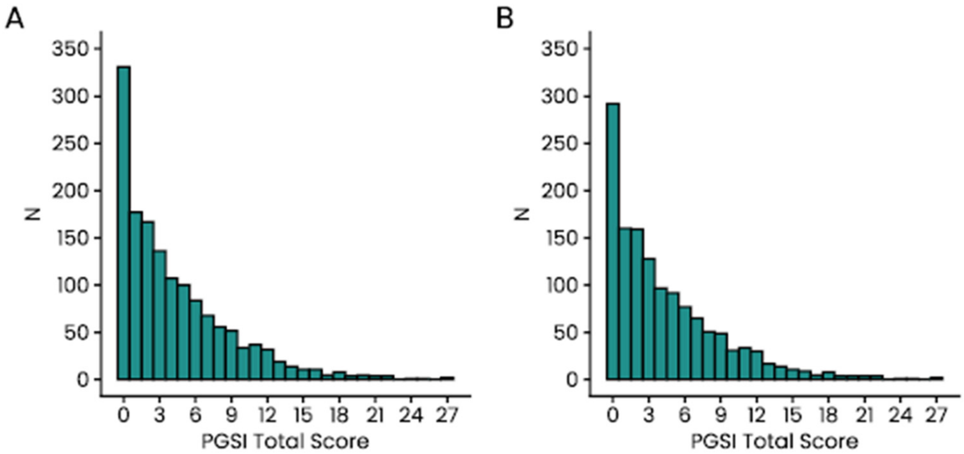

Fig. 1. PGSI total score distribution for the two samples
Note: ***Panel A:*** Total PGSI score distribution for the full sample (N = 1,470). ***Panel B:*** Total PGSI score distribution for the smaller subsample with six months of betting history (N = 1,349).

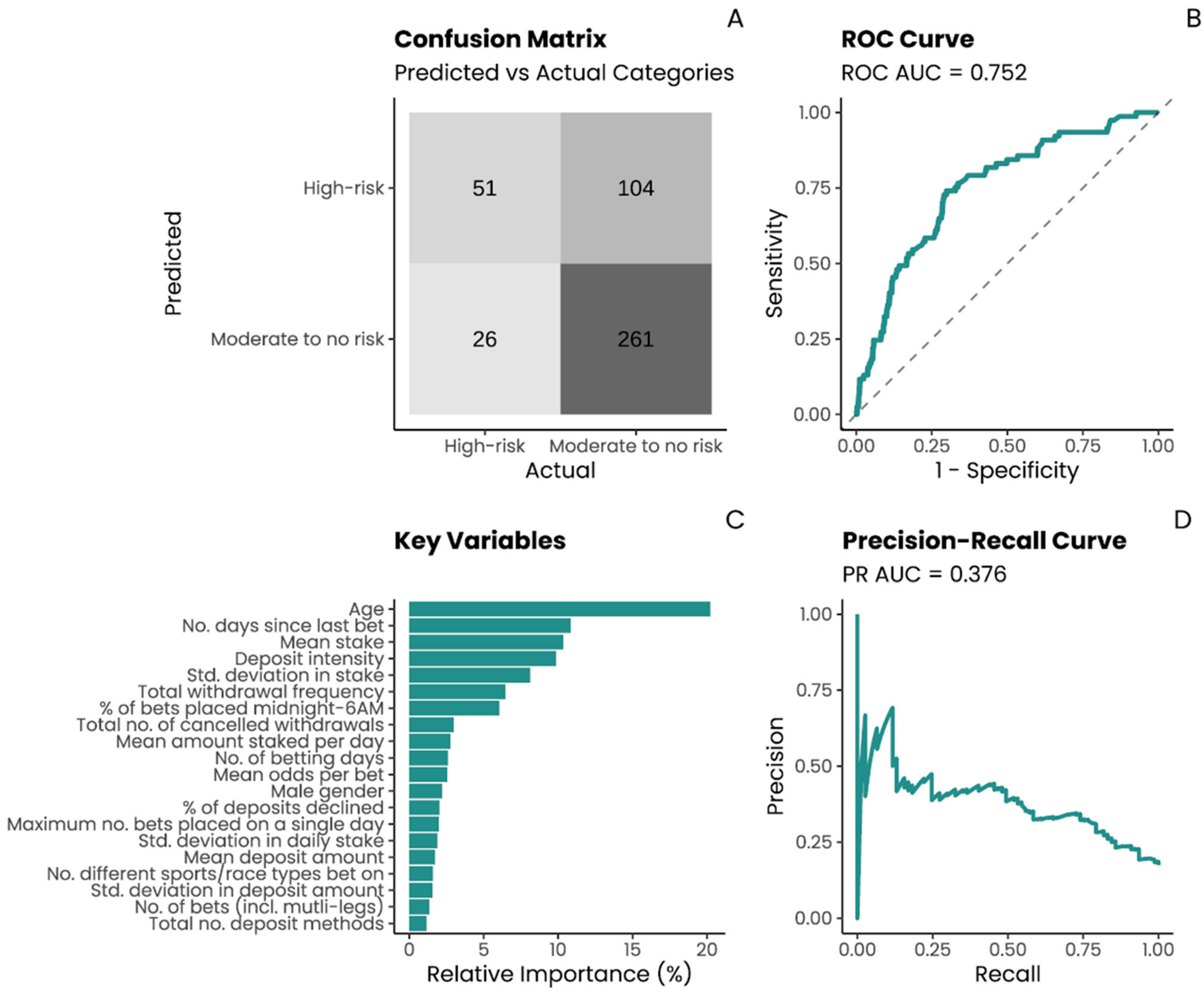

Fig. 2. Classification performance and variable importance for optimal Phase 1 model (30-day window using only account-based features)
Note: ***Panel A:*** Confusion matrix showing the model-predicted versus actual rates of consumers in each PGSI classification. ***Panel B:*** Receiver Operating Characteristic Curve showing the relationship between sensitivity (true positive rate) and 1-specificity (false positive rate) across decision cut-offs. ***Panel C:*** Top 20 most important variables used by the model to make its classifications. ***Panel D:*** Precision-recall curve showing the relationship between precision (i.e., proportion of those predicted to be a positive class who are truly at-risk) and recall (i.e., sensitivity, or true positive rate) across decision cut-offs.

all bets and transactions (i.e., deposits and withdrawals) made, as well as demographic details (i.e., age, sex, postcode, and account registration date), and engagement with consumer protection tools (CPTs, e.g., deposit limits, self-exclusion) for an extended timeframe covering two years before survey invitation (allowing assessment of CPTs set before our six-month pre-survey window).

We engineered multiple features (variables) summarising participants' gambling behaviour over the 30 days and six months preceding the date they completed the survey, which were used as predictors in our models. For example, for the 30-day pre-survey window, features such as total number of bets placed, total number of betting days, and deposit intensity (i.e., deposits per active betting day) were computed using each person's past-30-day data. The same features were computed for the six-month pre-survey window, with the addition of monthly average variables (e.g., average monthly net outcome, average monthly number of deposits) alongside six-month aggregate variables (e.g., overall net outcome, total number of bets placed). We aimed to generate features that achieved one or both of the following: [1] capture the general gambling profile of participants; and [2] reflect the features most predictive of risk status in similar studies (e.g., maximum number of bets placed in a single day, average number of deposits in a day; Murch et al., 2023; Perrot et al., 2022).

Income and related features: Pre-tax household income was assessed via a single question in the survey that presented income brackets to select (e.g., \$60,000–\$79,999 per year). We took the middle value within each bracket to represent annual income (e.g., "\$80,000–\$99,999/yr" became \$90,000).[2] Estimated income values were used to engineer features relating to participants' percentage of income spent on bets (i.e., staked) and deposited into their accounts.

Employment status: Assessed using a survey question with response options dichotomised into "Employed" (including responses indicating part-time and casual positions, full-time study, and retirement) and "Unemployed" (including responses indicating no current employment or being principally engaged in domestic duties).

Socio-economic status: We used customer postcodes provided by the operator to determine each person's Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) score. IRSAD scores are computed by the Australian Bureau of Statistics (ABS, 2023) and summarise information about the social and economic conditions of people and households in an area, including both relative advantage (e.g., % employed as professionals; % of high-income earners) and disadvantage (e.g., % unemployed; % of low-income earners). A low score indicates relatively greater disadvantage and a lack of advantage in general. As IRSAD scores are determined using postcodes, each participant may vary from this summary score.[3]

Other betting accounts: The survey included the following question: "*Including this account, how many online gambling accounts have you used in the past 12 months?*" (response options: 1, 2, 3, 4, and 5+). Responses were converted to a continuous variable (1–5) to reflect each participant's number of active accounts.

Estimations of gambling expenditure: Five questions asked participants about their past-30-day expenditure with their respective site, including net outcome (total amount won/lost), spend (total amount staked on bets), win (total amount won on bets), deposit, and withdrawal amounts (e.g., "*Approximately, how much money did you deposit with [site name] over the last 30 days?*"). We computed participants' actual past-30-day values to compare with their estimates. "Percentage discrepancy" variables were engineered to represent the difference between estimated and actual values as a percentage of the actual value (e.g., a participant who estimated that they deposited \$50 but actually deposited \$75 would have a percentage discrepancy value of 50%; $\left(\frac{75-50}{50}\right)\times 100$).

Life and gambling satisfaction: Modelled on a single item measure of life satisfaction from the Household, Income and Labour Dynamics in Australia survey (HILDA; Ambrey & Fleming, 2014) "*How satisfied are you with your life, all things considered?*", we developed a single question to assess gambling satisfaction: "*How satisfied are you with your gambling?*" (recorded on a scale from 0 [totally dissatisfied with my gambling] to 10 [totally satisfied with my gambling]).

Statistical analysis

Overall design. We employed a three-phase machine learning approach using the "tidymodels" (Kuhn & Wickham, 2020) framework in R (version 4.4.2; R Core Team, 2024). All analysis code is available here: https://github.com/Sydney-Informatics-Hub/Wagering-2025/. Our analysis utilised two distinct datasets filtered to keep only participants with complete survey data: a 30-day dataset comprising 1,470 participants and a six-month dataset containing 1,349 participants, with the difference reflecting the availability of extended historical data. Risk status was determined using the PGSI, with three classification schemes considered. First, a multiclass approach categorising participants into no, low, moderate, and high-risk groups (PGSI scores of 0, 1–2, 3–7, and 8+, respectively), consistent with conventional categorisations of the measure (Ferris & Wynne, 2001). Second, a binary classification with a PGSI

[2]To identify an appropriate value for participants who indicated that their income was in the highest bracket of more than \$200,000 a year, we used Australian Bureau of Statistics (2024) census data to look at the number of people in their income brackets for those earning above \$200,000 a year, computed the median value for each of these brackets, and worked out which value most closely split all people earning more than 200,000 a year evenly 50:50 (i.e., the median). This value was \$257,374 and was thus used as the estimated income value for this group.

[3]Whilst this variable was generated from existing account data, we consider it a "survey" feature in this study, as these variables are defined by the inclusion of additional information not routinely used for risk detection models.

≥5 threshold ("Moderate to high-risk" vs. "Low to no risk"). Third, a binary classification with a PGSI ≥8 threshold ("High-risk" vs. "Moderate to no risk"). Both binary classifications have been used in similar studies recently (Kairouz et al., 2023; Murch et al., 2023).

Model development. Data pre-processing followed a consistent protocol across all phases. The datasets were split into training and testing sets using a 70/30 ratio stratified to maintain class distribution of customers who had historically been flagged by their site's risk detection system and those who had not. Hyperparameters were tuned using 10-fold cross-validation repeated five times. Standard preprocessing steps included feature normalisation, categorical variable encoding, handling of missing values (median imputation for numeric variables and mode imputation for categorical variables), removal of near-zero variance predictors, and management of outliers. To address class imbalance in the outcome variable (where the minority to majority class ratio was less than 0.3), we implemented Synthetic Minority Oversampling Technique (SMOTE) to generate synthetic samples of the minority class. We confirmed that our models performed better (as measured by AUROC) on test data after being trained on data subjected to SMOTE than without its application.

Five distinct machine learning algorithms were considered: [1] Logistic Regression (using the "glm" library), [2] Random Forest ("ranger"), [3] XGBoost (extreme gradient boosting; "xgboost"), [4] Decision Tree ("rpart"), and [5] Neural Network ("nnet"). For model optimisation, we employed a space-filling design to generate 100 hyperparameter combinations.

Phased process of model testing and comparison. We used a phased approach to test and compare model performance under different scenarios. ***Phase 1*** considered the performance of the five model types for each of the three classification approaches using 31 features derived from 30-day account data, including behavioural metrics such as deposit frequency as well as demographic information (i.e., age and gender). ***Phase 2*** focused exclusively on the best-performing approach (i.e., model and classification type) from Phase 1, expanding to 62 features built from account history spanning six months before the survey, including both 30-day and six-month aggregate variables (e.g., number of betting days), as well as monthly averages (e.g., average monthly deposit amount). ***Phase 3*** further enhanced the feature sets by integrating survey data covering aspects such as gambling satisfaction and percentage of income spent (N = 11 [30-day]/13 [30-day + six-month]). The data window in all phases was relative to the date each participant completed the survey (i.e., 30 days or 6 months prior to the specific date of survey completion). A summary of feature sets included in models at each phase is presented in Table 2. A list of all features included in models is available in Supplemental Table S1, and feature definitions are presented in Supplemental Table S2.

Our primary measure of model performance was Area Under the Receiver Operating Characteristic curve (AUROC) values, which represents overall performance in a binary classification scenario. AUROC values typically range from 0.5 (no better than chance) to 1 (perfect discrimination). As this was our primary outcome measure, we computed 95% confidence intervals for these values using bootstrapping with 1,000 re-samples. We interpreted AUROC values using conventional guidelines: failed (0.50–0.59), poor (0.60–0.69), fair/adequate (0.7–0.79), good (0.80–0.89), or excellent (≥0.9) (Çorbacıoğlu & Aksel, 2023). The performance of the final models for each phase was also assessed using sensitivity, specificity, precision, accuracy, and balanced accuracy ([sensitivity + specificity]/2). Finally, we report Area Under the Precision-Recall Curve (AUPRC) values for our final models, consistent with the recommendations of Marionneau et al. (2025). AUPRC provides a single-number summary of binary classification performance, like AUROC, but focuses on the positive class and may therefore be less influenced by imbalanced datasets.

Ethics

Ethics approval for this study was obtained from the University of Sydney Ethics Committee (ID: 2023/029).

Table 2. Feature sets included in models

Feature set	Phase 1	Phase 2	Phase 3	
	30-day window: account data model	6-month window: account data model	30-day window: account & survey data model	6-month window: account & survey data model
Demographic characteristics	✓	✓	✓	✓
Past 30-day betting behaviour	✓	✓	✓	✓
Active use of CPTs	✓	✓	✓	✓
Past 6-month betting behaviour		✓		✓
Past 6-month changes to CPTs		✓		✓
Average monthly betting behaviour (past 6 months)		✓		✓
Survey response features			✓	✓
Total no. features	31	62	42	75

Note: CPT = Consumer protection tool (e.g., deposit limit).

All participants read a study information sheet and provided their consent before starting the survey.

RESULTS

Phase 1 models: establishing an optimal approach using 30 days of account data

We started by comparing the overall performance of the different model types for each of the three classification scenarios, optimising for Youden's J (J = sensitivity + specificity −1), and using the 30-day account data to develop these baseline models. The multiclass classification framework exhibited limited discriminative capacity, with even the highest-performing model (decision tree) achieving only a test AUROC of 0.613 (Table 3). In contrast, binary classification frameworks consistently demonstrated enhanced discriminative capability, with particularly strong results for the PGSI ≥8 threshold. The binary PGSI ≥8 or lower classification approach implemented through XGBoost achieved the highest test AUROC (0.752). Accordingly, more extensive outcomes for this model are presented in Fig. 2. This model correctly identified 66.2% of high-risk cases (sensitivity) and 71.5% of moderate-to-no-risk cases (specificity).

The XGBoost algorithm provided the best discriminative capacity for identifying high-risk gambling behaviour while maintaining an appropriate balance between false positives and false negatives (see Supplemental Table S3 for model parameters when using alternative optimisation metrics for model selection other than Youden's J index; e.g., AUROC, sensitivity). Based on these initial assessments, we established the PGSI ≥8 threshold using XGBoost as our preferred model architecture for subsequent phases.

Phase 2 model: expanding to six months of account data

Changing the temporal framework to the preceding six months yielded a marginal reduction in overall performance (test AUROC of 0.743) but slight improvements in sensitivity (67.5%) and specificity (72.1%). Figure 3 presents the full set of results for this model.

Phase 3 models: integration of survey responses

Integration of survey data in Phase 3 substantially enhanced overall performance, particularly for the 30-day model (Fig. 4), which achieved the highest overall test AUROC of 0.850. Sensitivity remained at 66.2% for the 30-day model with added survey features, whereas specificity improved to 84.4%, representing a 12.9% increase over the 30-day model in Phase 1. The six-month model with survey features also demonstrated improved overall performance relative to the six-month model in Phase 2, with a test AUROC of 0.832 (Fig. 5). Gains were seen in specificity, which increased by 12.0 to 84.1%, while sensitivity decreased by 1.2 to 66.3%.

Variable importance

In Phases 1 and 2, age (20.2% relative importance in the optimal 30-day model; 9.5% in the six-month model), deposit intensity (9.9%; 29.3%), average stake size (10.3%; 11.2%), and the number of days since placing a bet (10.8%;

Table 3. Comparison of classification approaches

	Cross-validation metrics		Test set performance		Model stability	
Model	AUROC	Std. Error	AUROC	Accuracy	ROC Δ	ROC stability
Classification: multi-class PGSI groupings						
Decision tree	0.628	0.005	0.613	0.317	0.014	0.008
XGBoost	0.663	0.004	0.610	0.305	0.054	0.007
Multinomial regression	0.643	0.004	0.601	0.333	0.042	0.007
Random Forest	0.656	0.005	0.595	0.319	0.061	0.007
Neural network	0.611	0.005	0.570	0.290	0.042	0.008
Classification: PGSI ≥5						
Logistic regression	0.701	0.006	0.678	0.688	0.023	0.009
Random Forest	0.723	0.007	0.669	0.683	0.054	0.009
XGBoost	0.707	0.007	0.652	0.686	0.055	0.009
Neural network	0.653	0.007	0.647	0.654	0.006	0.011
Decision tree	0.652	0.008	0.639	0.672	0.013	0.012
Classification: PGSI ≥8						
XGBoost	**0.719**	**0.008**	**0.752**	**0.706**	**0.033**	**0.011**
Random Forest	0.700	0.008	0.705	0.756	0.006	0.011
Decision tree	0.652	0.010	0.704	0.686	0.052	0.016
Logistic regression	0.673	0.008	0.697	0.654	0.023	0.012
Neural network	0.645	0.009	0.650	0.670	0.005	0.014

Note: AUROC = Area Under the Receiver Operating Characteristic Curve; Std Error = standard error of the mean AUROC achieved in cross-validation folds; Accuracy = percentage of total cases correctly classified; ROC Δ: Absolute difference between cross-validation and test set AUROC; ROC stability = Stability of ROC across folds. The emboldened row represents the best-performing model.

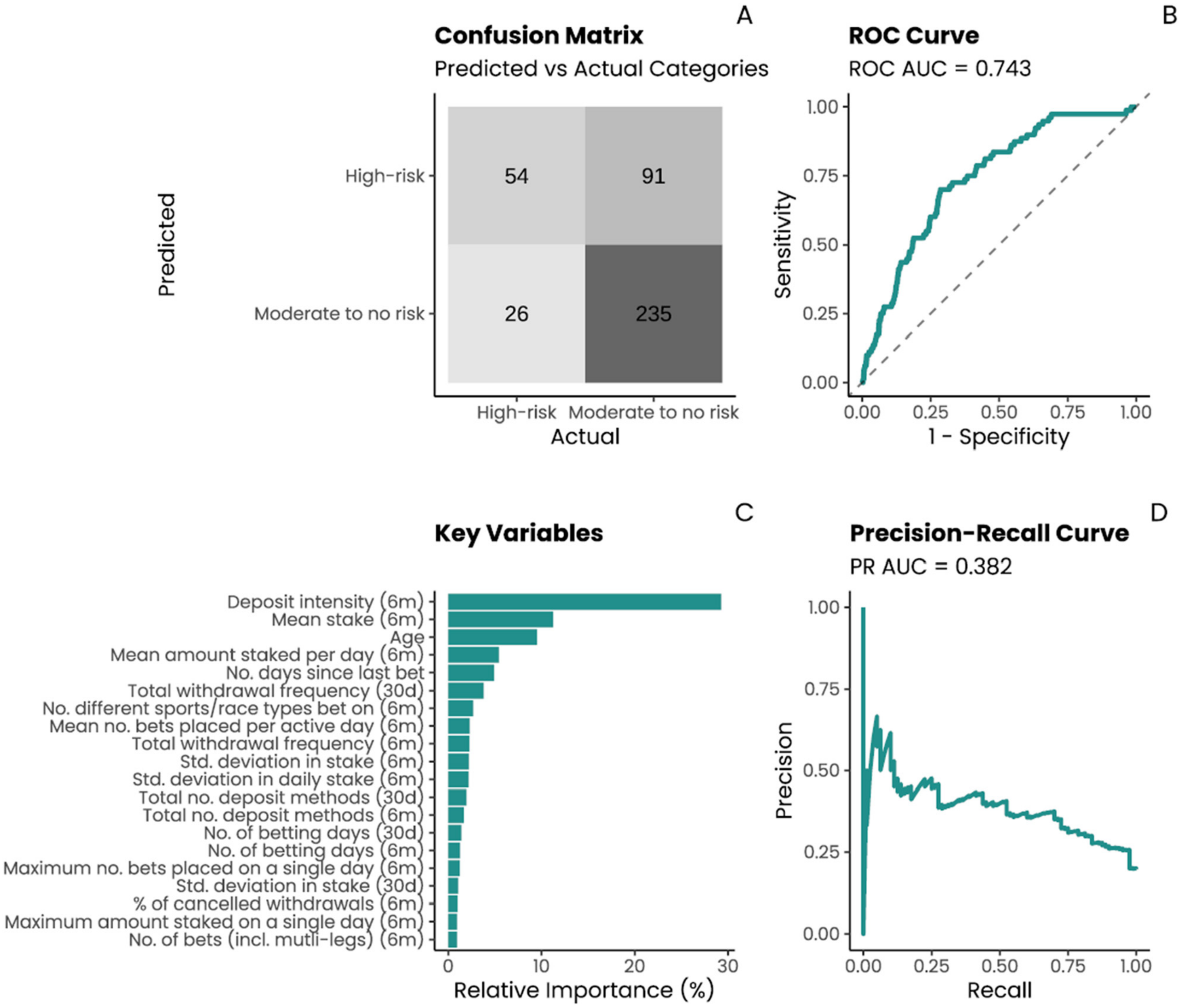

Fig. 3. Classification performance and variable importance for Phase 2 model (six-month window using only account-based features)
Note: ***Panel A:*** Confusion matrix showing the model-predicted versus actual rates of consumers in each PGSI classification. ***Panel B:*** Receiver Operating Characteristic Curve showing the relationship between sensitivity (true positive rate) and 1-specificity (false positive rate) across decision cut-offs. ***Panel C:*** Top 20 most important variables used by the model to make its classifications. ***Panel D:*** Precision-recall curve showing the relationship between precision (i.e., proportion of those predicted to be a positive class who are truly at-risk) and recall (i.e., sensitivity, or true positive rate) across decision cut-offs. "6m" = feature based on six-month window preceding participation; "30d" = feature based on 30-day window preceding participation.

4.9%) served as key predictors (Figs 2C and 3C). In Phase 3, gambling satisfaction emerged as the primary predictor (32.2% in the 30-day model; 40.2% in the six-month model) in both temporal frameworks (Figs 4C and 5C). Other top predictors in Phase 3 models included the number of active gambling accounts (12.0%; 9.0%) and the difference between estimated and actual withdrawal (4.5%; 7.9%), deposit (4.6%; 2.3%), and net outcome (3.2%; 2.6%) amounts.

Phase 4 models: incremental gains from select survey features

Given the predominance of two self-report variables in Phase 3 models—gambling satisfaction and number of active gambling accounts—we conducted a fourth exploratory analysis phase, building a series of models that systematically integrated or excluded these features to determine their relative value in improving classification performance, plotting ROC and precision-recall curves for each model (Fig. 6). AUROC was highest when both survey variables were incorporated into a base model with account features from the 30-day history, and AUPRC was highest when only gambling satisfaction was added (see Supplemental Figures S2–4 for full model outcomes).

Overall model comparison

Table 4 compares the performance metrics of models from all four phases, demonstrating superior performance among the Phase 3 and 4 models that combined account data with varying numbers of survey features.

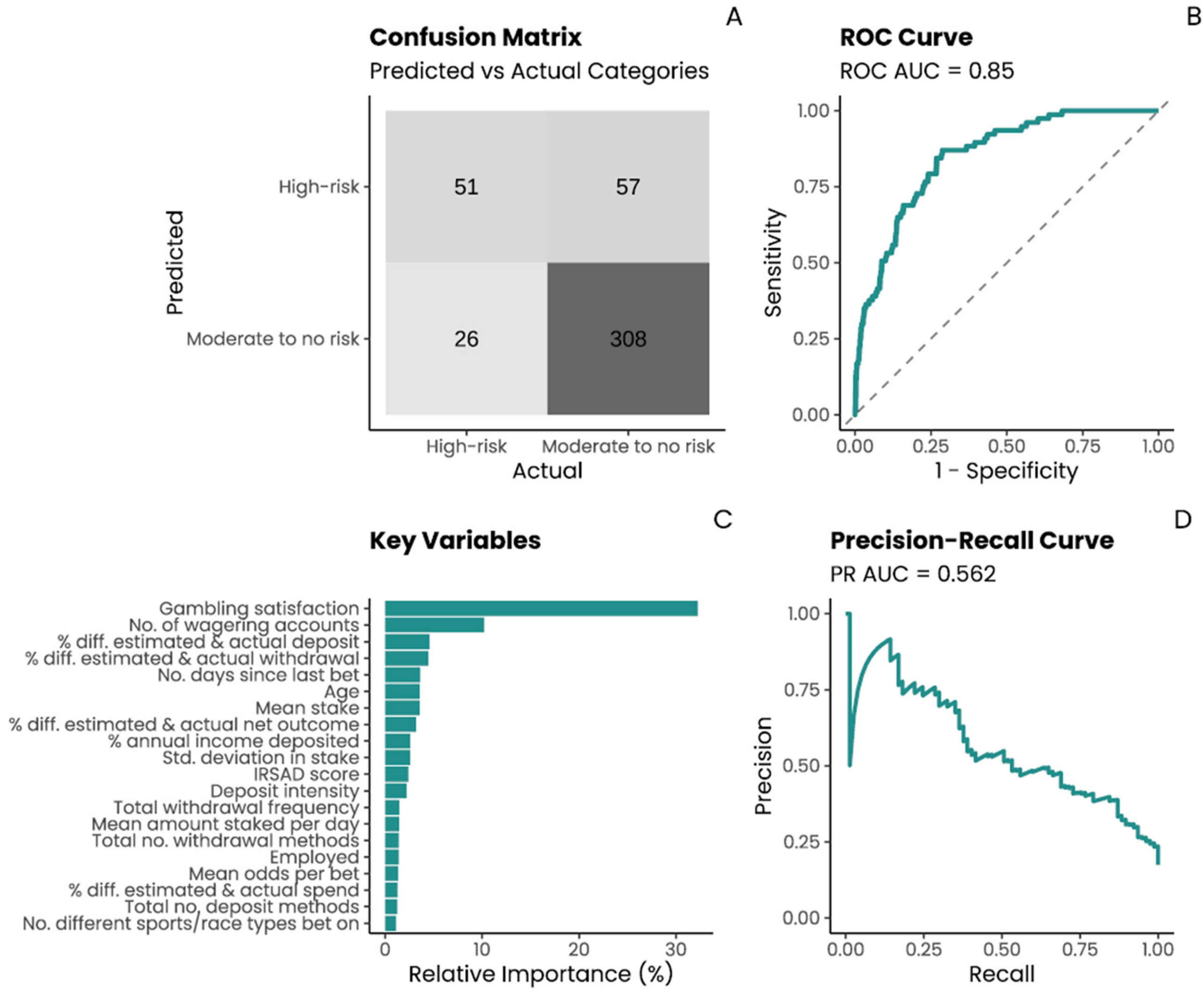

Fig. 4. Classification performance and variable importance for Phase 3 model (30-day window combining account- and survey-based features)

Note: ***Panel A:*** Confusion matrix showing the model-predicted versus actual rates of consumers in each PGSI classification. ***Panel B:*** Receiver Operating Characteristic Curve showing the relationship between sensitivity (true positive rate) and 1-specificity (false positive rate) across decision cut-offs. ***Panel C:*** Top 20 most important variables used by the model to make its classifications. ***Panel D:*** Precision-recall curve showing the relationship between precision (i.e., proportion of those predicted to be a positive class who are truly at-risk) and recall (i.e., sensitivity, or true positive rate) across decision cut-offs.

DISCUSSION

Our machine learning models achieved adequate to good performance in correctly classifying online sports and race betting customers into higher- and lower-risk PGSI categories (see Table 5 for a lay-summary of findings). The models we built showed reasonable sensitivity (66.2–75.6) and specificity (71.5–84.4), but poor precision (32.9–50.5). This is similar to previous studies that have used a comparable design to ours (precision rates: 29.5–49.6; Kairouz et al., 2023; Murch et al., 2023). In this study, precision refers to the proportion of customers predicted to be high-risk who reported PGSI scores ≥8. Thus, a large proportion of those flagged by these models for intervention may not be truly high-risk, although this may be a necessary trade-off to ensure adequate sensitivity (i.e., detection of high-risk customers; Murch et al., 2023).

Model performance was stronger when predicting PGSI scores of ≥8 compared to scores of ≥5, consistent with the ≥8 threshold reflecting a more distinct and severe symptom profile (Currie, Hodgins, & Casey, 2013). Kairouz et al. (2023) found marginally improved model performance for the ≥8 over the ≥5 classification, although Murch et al. (2023) found the converse. Overall, our findings and those of previous studies (Kairouz et al., 2023; Murch et al., 2023) indicate that these models can perform adequately across different risk thresholds. In applied settings, the choice of how to define the high-risk group may be better guided by risk detection preferences, rather than model performance optimisation. Selecting a lower PGSI cut-off may be

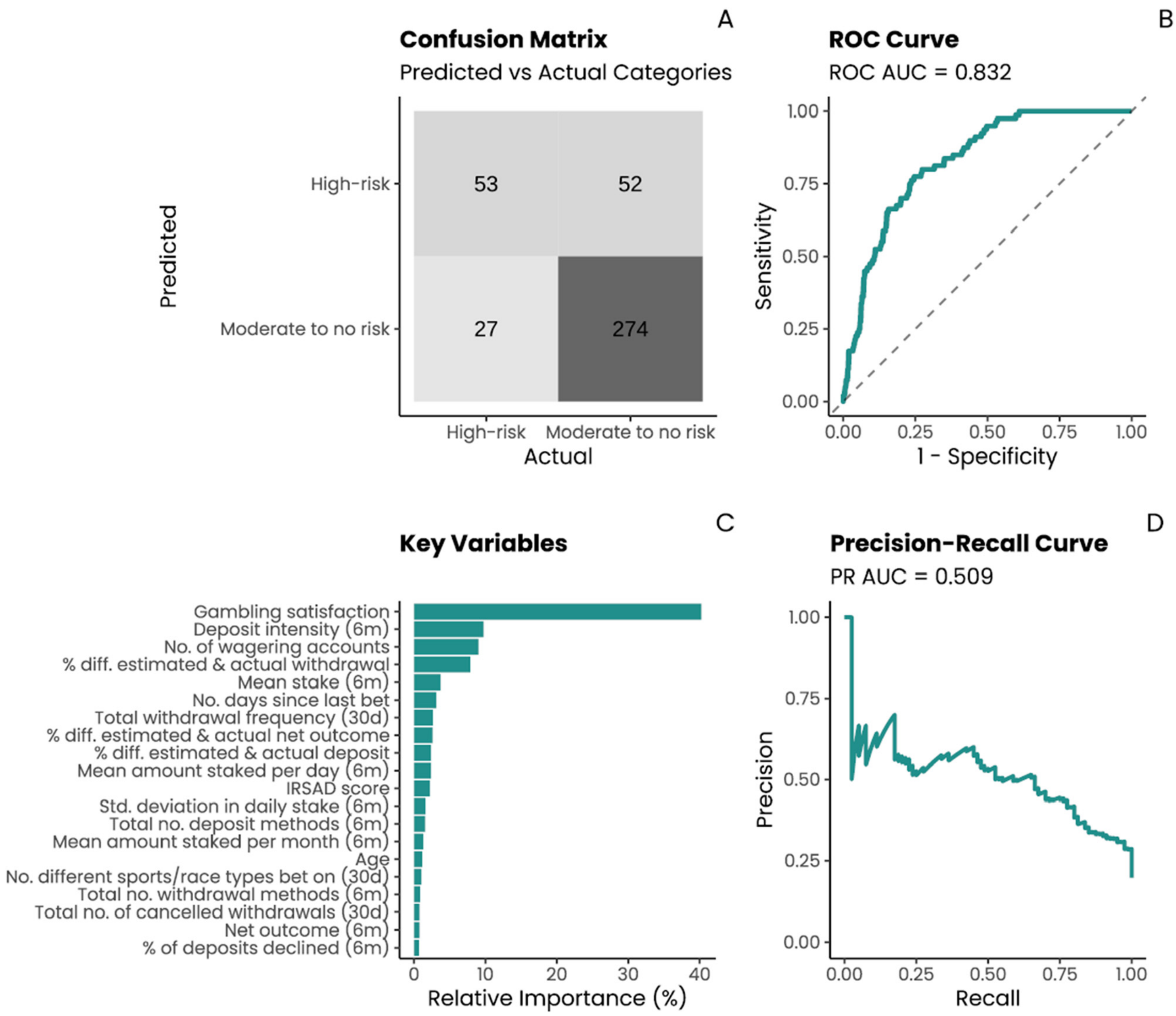

Fig. 5. Classification performance and variable importance for Phase 3 model (six-month window combining account- and survey-based features)

Note: ***Panel A:*** Confusion matrix showing the model-predicted versus actual rates of consumers in each PGSI classification. ***Panel B:*** Receiver Operating Characteristic Curve showing the relationship between sensitivity (true positive rate) and 1-specificity (false positive rate) across decision cut-offs. ***Panel C:*** Top 20 most important variables used by the model to make its classifications. ***Panel D:*** Precision-recall curve showing the relationship between precision (i.e., proportion of those predicted to be a positive class who are truly at-risk) and recall (i.e., sensitivity, or true positive rate) across decision cut-offs. 6m = variable based on six-month window preceding participation; 30d = variable based on 30-day window preceding participation.

preferred when emphasising the early detection of potentially at-risk individuals, whilst using a higher cut-off may reflect a preference for reducing misclassifications of those not experiencing problems.

Among the models we tested, XGBoost achieved the best classification accuracy. To our knowledge, our study is the first to include XGBoost when predicting self-reported problem gambling (but not VSE), with prior studies finding Random Forest performs best in this context (Auer & Griffiths, 2022; Hopfgartner et al., 2024; Kairouz et al., 2023; Murch et al., 2023; Perrot et al., 2022). Outside of the gambling field, XGBoost models have been found to achieve good to excellent accuracy in different classification scenarios (Bentéjac, Csörgő, & Martínez-Muñoz, 2021). Gambling researchers should consider XGBoost and other gradient-boosted models in future risk detection studies.

Adequate classification was achieved using only 30 days of data, and longer account histories (six months) did not meaningfully improve model performance. This is consistent with Andersson et al. (2025), who found models trained on 30, 60, and 90 days of data performed similarly. The 30-day models in our study appeared to provide largely similar sensitivity and specificity, but lower precision (i.e., lower proportions of those predicted to be at-risk who are truly at-risk) than six-month models. It is important to note that the 30-day data used in this study was not necessarily the first 30

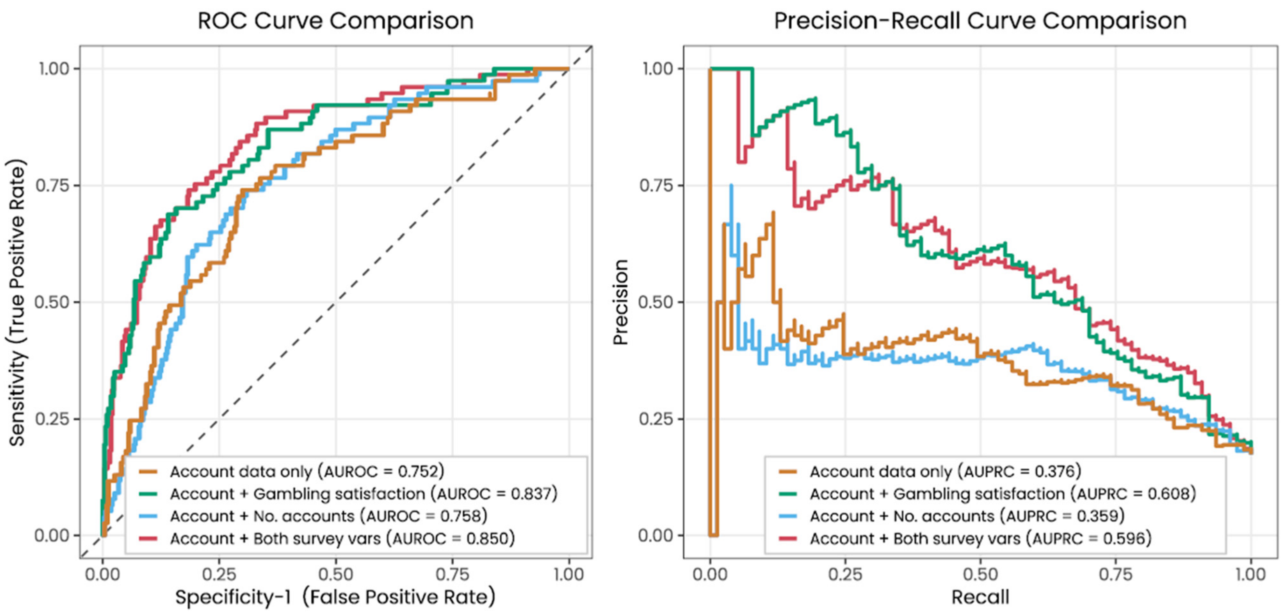

Fig. 6. Comparison of models including or excluding highly predictive survey variables (Phase 4)

Note: Outcomes for four separate models are presented in both panels, including a baseline model using only past-30-day account data, the baseline model plus the gambling satisfaction variable, the baseline model plus the number of accounts variable, and the baseline model plus both survey variables. ***Panel A:*** Receiver Operating Characteristic Curve plot showing the relationship between sensitivity (true positive rate) and 1-specificity (false positive rate) across decision cut-offs for each model. ***Panel B:*** Precision-recall curve plot showing the relationship between precision (i.e., proportion of those predicted to be in the positive class who are truly at-risk) and recall (i.e., sensitivity, or true positive rate) across decision cut-offs for each model.

Table 4. Model performance comparison

Model	AUROC [95% CIs]	AUPRC	Sensitivity	Specificity	Precision	Accuracy	Balanced accuracy
Phase 1							
30 days - account features	0.752 [0.692, 0.808]	0.376	66.23	71.51	32.90	70.59	68.87
Phase 2							
6 months - account features	0.743 [0.690, 0.796]	0.382	67.50	72.09	37.24	71.18	69.80
Phase 3							
30 days - account & survey features	**0.850 [0.805, 0.888]**	0.562	66.22	**84.38**	47.22	**81.22**	75.31
6 months - account & survey features	0.832 [0.790, 0.874]	0.509	66.25	84.05	**50.48**	80.54	75.15
Phase 4							
30 days - account & gambling satisfaction	0.837 [0.781, 0.883]	**0.608**	72.73	77.53	40.60	76.70	75.13
30 days - account & no. accounts	0.758 [0.704, 0.815]	0.359	64.94	74.79	35.21	73.08	69.87
30 days - account & both survey vars	**0.850 [0.802, 0.897]**	0.596	**75.64**	80.27	45.04	79.46	**77.96**

Note: AUROC = Area under the receiver operating characteristic curve; CIs = Confidence Intervals; AUPRC = Area under the precision recall curve; Sensitivity = Percentage of at-risk participants predicted to be at-risk by the model (true positive rate); Specificity = Percentage of not-at-risk participants predicted to be not-at-risk by the model (true negative rate); Precision = Percentage of participants predicted to be at-risk by the model who were truly at-risk; Accuracy = Total percentage of all participants correctly classified; Balanced accuracy = Mean of sensitivity and specificity; Emboldened values = highest rate achieved.

days of someone's account history; nonetheless, our findings indicate risk-detection efforts can begin early after registration. Future research should determine whether models built using data from the first 30 days after account registration, or even shorter periods, can achieve adequate performance.

Table 5. Summary of key findings

- Machine learning models are computer-based algorithms that can analyse large amounts of data to find patterns and make predictions.
- We aimed to determine whether these models could be used to identify online sports and race betting customers with self-reported gambling problems (PGSI scores ≥5 and ≥8) using behavioural tracking data that is routinely collected by betting sites.
- The models distinguished between customer risk groups better when we used PGSI scores ≥8 to define the high-risk group (compared to when using scores ≥5).
- Accuracy: These models were able to correctly classify 71-81% of all customers according to their actual status as either no-to-moderate-risk (PGSI <8) or high-risk (PGSI ≥8).
- Sensitivity: 65–76% of high-risk customers were correctly classified as high-risk by the models we tested, meaning 24-35% were incorrectly classified as no-to-moderate-risk customers.
- Specificity: 71–84% of no-to-moderate-risk customers were correctly classified as such by the models we tested, meaning 16–29% were incorrectly classified as high-risk customers.
- Precision: only 33–50% of those predicted to be high-risk by our models were truly high-risk (i.e., PGSI ≥8).
- We built models using either 30 days or six months of customers' behavioural account data and found they performed similarly, suggesting extended periods of historical data are not needed to use these models and they can be implemented soon after registration.
- The account-based variables most predictive of risk status were young age, high deposit intensity (defined as the average number of deposits per active betting date), high average stake size, and fewer days since last placing a bet.
- When we integrated responses from a customer survey, the models' ability to accurately classify customers substantially improved.
- The self-report variables most predictive of risk status were gambling satisfaction and the number of online gambling accounts someone held.
- The best model we built (see Figure 4) combined 30 days of behavioural account data with self-report variables.

One of the most novel findings from this study was the substantial performance improvements achieved by integrating survey variables into models. Our best performing model combined 30-day account data and survey variables, suggesting self-reports and recent behaviour may act synergistically to predict problem gambling. We observed meaningful improvements by adding just two survey variables to models in isolation or combination: gambling satisfaction and the number of active wagering accounts. Gambling satisfaction was strongly predictive of problem gambling risk status, with lower scores observed among higher-risk customers. This novel variable is likely less biased than directly asking customers about their experience of gambling harms as people may deny and/or fail to recognise these. We recommend further research to explore the concept of gambling satisfaction and its value as an indicator of risk.

There are several reasons why self-reported data may be more predictive of risk than account-based variables. First, the relationship between time or money spent gambling and the perceived sense of problems may be complex and variable. There is likely variation between individuals in [1] the number of internal and external resources available to help them manage the impact of their gambling, and [2] their interpretation of the impact gambling is having on them. Second, account-based variables only capture behaviour on one site. Survey-based variables, by comparison, can reflect a person's entire experience with gambling. This explanation is supported by our finding that the number of active online gambling accounts someone held provided a strong signal for differentiating risk groups. A single customer view of gambling using bank or digital wallet records could provide a complete picture of consumption across sites and has demonstrated value in a UK-based study of open banking data (Zendle & Newall, 2024). A dedicated service that can track individuals' transactions across gambling sites irrespective of their payment method is likely to be most effective at identifying and supporting high-risk consumers; this could be developed by regulators, financial institutions, or trusted third parties using account data, ideally in real-time (Swanton, Gainsbury, & Blaszczynski, 2019).

Finally, because both the survey predictors and PGSI scores rely on self-report, any individual differences in response styles will influence both sets of measures in a correlated way. This shared method variance can artificially inflate the association between survey predictors and the outcome, a problem referred to as "common method bias" (Kock, Berbekova, & Assaf, 2021). A second and related reason is that the timing of collection was identical, meaning participants' current emotional state and recent gambling experiences could affect all their survey responses simultaneously. For example, dissatisfaction with recent gambling outcomes might lead someone to endorse more PGSI items and report lower gambling satisfaction. These issues highlight the need for future research to confirm the predictive ability of key self-report variables identified here with other possible measures of risk (e.g., reports from operator contact with customers) and to separate the timing of predictor and outcome measurement.

Our findings should be viewed in the context of some limitations. We considered account data from only two sites and were unable to track gambling beyond the single site each customer was registered with. As our outcomes demonstrate, high-risk individuals are more likely to gamble with multiple sites (see also: The UK Behavioural Insights Team, 2021). We attempted to address this by capturing the number of other sites recently gambled with, but this is unlikely to fully reflect the extent and variation of gambling across other sites. Second, whilst we obtained a moderately large sample (N = 1,349–1,470), similar studies in the field have obtained sample sizes ranging from 5,404 (Perrot et al., 2022) to 14,261 (Luquiens et al., 2016), potentially enhancing the representativeness of their datasets. However, unlike prior studies, our survey was extensive and burdensome for participants, including the PGSI alongside measures of estimated expenditure, demographic characteristics, gambling satisfaction, and several other constructs not central to the present analyses. Sample size concerns in this type of research highlight a limitation of using self-reported problem gambling scores as the target variable. While VSE has limitations in this context that are discussed above (e.g., heterogeneous motivations for self-excluding), it is routinely available in operator data without the need to encourage customers to complete a survey. Sample representativeness is another concern. In a separate analysis of the dataset used here, our team was able to determine that the sample used in this study was not fully representative of the wider customer base, as more engaged bettors self-selected into the survey (Heirene et al., 2025). This is unsurprising, as more engaged bettors are more likely to be active and interested in completing a survey about their gambling. Future research is needed to understand how to scale the collection of self-report measures on online sites and ensure representation across activity levels.

Practical implications and conclusions

Our findings suggest that machine learning models can be used to identify online bettors who may be experiencing gambling problems. They correctly identify most customers who report gambling problems as high-risk, although a notable proportion of those flagged as being high-risk may not be experiencing problems. False positives are a common feature of these models and may be a trade-off required to identify most truly high-risk customers. Operators can employ these models with just 30 days of historical customer data, using them to inform responsive interventions. Self-report data that are not personal, sensitive, or difficult to provide accurately can substantially enhance model performance. Collecting information such as customers' self-reported gambling satisfaction and number of other accounts can shift towards more person-centred approaches to risk detection that account for individual experiences and behaviours across platforms.

Funding sources: The customer data analysed for this project was provided by the online wagering operator Entain, who also facilitated the collection of survey data (PI: Prof. Gainsbury, University of Sydney, IRMA ID: 217448 Safer Gambling Project). Entain does not and will not have access to survey responses. This research was funded in part by a grant from the International Center for Responsible Gaming (ICRG), funded by Bally's Corporation. Its contents are solely the responsibility of the author(s) and do not necessarily represent the official views of the ICRG or Bally's Corporation.

RH was supported by a post-doctoral fellowship from the New South Wales (NSW) Responsible Gambling Fund as part of the Gambling Research Capacity Grant program, administered by the Office of Responsible Gambling (ORG).

Entain, ICRG, Bally's Corporation, and the NSW ORG did not have any input into the research question(s), design, data analysis, or interpretation of findings of this study. The first version of this article was provided to Entain in advance of submission to the journal, but Entain was unable to request changes to this work or preclude dissemination of any findings.

Authors' contribution: RMH: conceptualization (lead), data curation (lead), formal analysis, funding acquisition, investigation (lead), methodology (lead), project administration, visualisation, writing - original draft (lead), and writing - review & editing (lead). EZ: formal analysis (lead), visualisation (lead), writing - original draft, and writing - review & editing. DV: conceptualisation, formal analysis, and writing - review & editing. CTdL: formal analysis, writing - original draft, and writing - review & editing. EH: writing - original draft, and writing - review & editing. SMG: conceptualisation, data curation, funding acquisition (lead), investigation, methodology, supervision, and writing - review & editing.

Conflict of interest: RH has worked on a project funded by Responsible Wagering Australia (a representative body of Australian online wagering operators; University of Sydney, 2019–2021) and as an independent, sub-contracted statistical consultant for PRET Solutions Inc on a commissioned project (funded by the Australian Casino operator Crown; 2023). RH has received funding from the International Center for Responsible Gaming (ICRG); the Brain and Mind Centre and wider University of Sydney; and the New South Wales (NSW) Responsible Gambling Fund, administered by the Office of Responsible Gambling (ORG).

EZ and DV have no competing interests to declare.

CTdL has received funding from ZonMW, a legal entity created by the Dutch Government that funds healthcare research. He has not received any funding to support his involvement in this research.

ELYH has worked on projects funded through the ICRG, awarded to SG and RH.

SG has received direct and indirect funding since 2020 through the University of Sydney from Australian Leisure and Hospitality Group Pty Ltd, Entain Australia, Sportsbet, NSW Office of Responsible Gambling, West HQ, Brain and Mind Centre, Cambridge Health Alliance, Aristocrat Leisure Limited, and ICRG. SG has received consulting funds for providing subject matter expertise for New Zealand Crown Counsel, NSW Liquor & Gaming Authority, New Zealand Gaming Machine Association, Singapore Ministry of Health, Betcloud, NZ BlueCloud, UK Behavioural Insights Team, KPMG, QBE, Coms Systems Limited, Advance Gaming (NZ) Limited, GambleAware, Star Entertainment, GREO, Senet, and Norths Collective. SG has received honorarium and/or travel costs for presentations for Cyprus National Betting Authority, Asian Racing Federation, Leagues Club Australia, Australian Cricketers Association, Star Entertainment, CAMH, Behavioural Insights Team, National

Council on Problem Gambling, GambleAware, GREO, Informa, and Washington State Council, European Lotteries Association. SG holds unpaid appointments as an invited member on the NSW Independent Panel on Gambling Reform, is the Pillar Champion for Technology and Environment for the QLD Responsible Gambling Advisory Committee, is a board member for the Asian Racing Federation Council on Anti-Illegal Betting and Related Financial Crime. SG receives an honorarium for her role as Co-Editor-in-Chief for International Gambling Studies from Taylor & Francis.

Data availability: Data used for this study are commercially sensitive and cannot be shared beyond the research team. The survey and additional methodological details can be accessed here: https://osf.io/vdsmw. All analysis code has been made openly available here: https://github.com/Sydney-Informatics-Hub/Wagering-2025/

Acknowledgements: The authors acknowledge the technical assistance provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1556/2006.2025.00525.

REFERENCES

ACMA. (2022). *Online gambling in Australia, findings from the 2021 ACMA annual consumer survey*. Australian Communications and Media Authority. https://www.acma.gov.au/publications/2022-02/report/online-gambling-australia-snapshot.

Allami, Y., Hodgins, D. C., Young, M., Brunelle, N., Currie, S., Dufour, M., … Nadeau, L. (2021). A meta-analysis of problem gambling risk factors in the general adult population. *Addiction (Abingdon, England)*, *116*(11), 2968–2977. https://doi.org/10.1111/add.15449

Ambrey, C. L., & Fleming, C. M. (2014). Life satisfaction in Australia: Evidence from ten years of the HILDA survey. *Social Indicators Research*, *115*(2), 691–714. https://doi.org/10.1007/s11205-012-0228-0

Andersson, S., Carlbring, P., Lyon, K., Bermell, M., & Lindner, P. (2025). Insights into the temporal dynamics of identifying problem gambling on an online casino: A machine learning study on routinely collected individual account data. *Journal of Behavioral Addictions*, *14*(1), 490–500. https://doi.org/10.1556/2006.2025.00013

Auer, M., & Griffiths, M. D. (2014). Personalised feedback in the promotion of responsible gambling: A brief overview. *Responsible Gambling Review*, *1*(1), 27–36.

Auer, M., & Griffiths, M. D. (2022). Using artificial intelligence algorithms to predict self-reported problem gambling with account-based player data in an online casino setting. *Journal of Gambling Studies*, *39*(3), 1273–1294. https://doi.org/10.1007/s10899-022-10139-1

Auer, M., & Griffiths, M. D. (2023). Reasons for gambling and problem gambling among Norwegian horse bettors: A real-world study utilizing combining survey data and behavioral player data. *International Journal of Mental Health and Addiction*, *21*(2), 740–755. https://doi.org/10.1007/s11469-020-00442-6

Auer, M., & Griffiths, M. D. (2026). Differences in online gambling expenditure between players from Germany, Spain, Netherlands, Great Britain, US and Canada: A largescale online player tracking study. *Acta Psychologica*, *266*, 106864. https://doi.org/10.1016/j.actpsy.2026.106864

Australian Bureau of Statistics. (2023). *Socio-Economic Indexes for Areas (SEIFA), Australia*. ABS. https://www.abs.gov.au/statistics/people/people-and-communities/socio-economic-indexes-areas-seifa-australia/2021.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*(3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Bowden-Jones, H., Hook, R. W., Grant, J. E., Ioannidis, K., Corazza, O., Fineberg, N. A., … Chamberlain, S. R. (2022). Gambling disorder in the UK: Key research priorities and the urgent need for independent research funding. *The Lancet Psychiatry*, *9*(4), 321–329. https://doi.org/10.1016/S2215-0366(21)00356-4

Catania, M., & Griffiths, M. D. (2021). Understanding online voluntary self-exclusion in gambling: An empirical study using account-based behavioral tracking data. *International Journal of Environmental Research and Public Health*, *18*(4), 4. https://doi.org/10.3390/ijerph18042000

Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*. https://doi.org/10.4103/tjem.tjem_182_23

Currie, S. R., Hodgins, D. C., & Casey, D. M. (2013). Validity of the problem gambling severity index interpretive categories. *Journal of Gambling Studies*, *29*(2), 311–327. https://doi.org/10.1007/s10899-012-9300-6

Delfabbro, P., Parke, J., & Catania, M. (2023). Behavioural tracking and profiling studies involving objective data derived from online operators: A review of the evidence. *Journal of Gambling Studies*, *40*(2), 639–671. https://doi.org/10.1007/s10899-023-10247-6

European Gaming & Betting Association. (2025). *Gambling market revenue and online share by country (2023)*. https://www.egba.eu/resource-post/online-gambling-shares-of-national-gambling-markets-in-europe/.

Ferris, J. A., & Wynne, H. J. (2001). *The Canadian problem gambling index: User manual*. Canadian Centre on Substance Abuse.

Finkenwirth, S., MacDonald, K., Deng, X., Lesch, T., & Clark, L. (2020). Using machine learning to predict self-exclusion status in online gamblers on the playNow.com platform in British Columbia. *International Gambling Studies*, 1–18. https://doi.org/10.1080/14459795.2020.1832132

Forrest, D., & McHale, G. (2022). *Patterns of play. Technical report 2: Account data stage*. NatCen Social Research.

Gainsbury, S. M. (2011). Player account-based gambling: Potentials for behaviour-based research methodologies. *International Gambling Studies*, *11*(2), 153–171. https://doi.org/10.1080/14459795.2011.571217

Gainsbury, S. M. (2014). Review of self-exclusion from gambling venues as an intervention for problem gambling. *Journal of Gambling Studies*, *30*(2), 229–251. https://doi.org/10.1007/s10899-013-9362-0

Ghaharian, K., Abarbanel, B., Phung, D., Puranik, P., Kraus, S., Feldman, A., & Bernhard, B. (2023). Applications of data science for responsible gambling: A scoping review. *International Gambling Studies*, *23*(2), 289–312. https://doi.org/10.1080/14459795.2022.2135753

Greer, N., Jenkinson, R., Vandenberg, B., & Sakata, K. (2023). *Regular online betting in Australia, 2022: National Gambling Trends Study*. Australian Gambling Research Centre, Australian Institute of Family Studies. https://aifs.gov.au/all-research/research-snapshots/regular-online-betting-australia-2022.

Griffiths, M. D., & Auer, M. (2016). Should voluntary self exclusion by gamblers be used as a proxy measure for problem gambling? *MOJ Addiction Medicine & Therapy*, *2*(2), 1–3. https://doi.org/10.15406/mojamt.2016.2.00019

Hayer, T., & Meyer, G. (2011). Internet self-exclusion: Characteristics of self-excluded gamblers and preliminary evidence for its effectiveness. *International Journal of Mental Health and Addiction*, *9*(3), 296–307. https://doi.org/10.1007/s11469-010-9288-z

Heirene, R. M., Cobb-Clark, D., Tymula, A., Santos, T., & Gainsbury, S. M. (2025). Non-response bias in gambling surveys. *International Gambling Studies*, 1–24. https://doi.org/10.1080/14459795.2025.2530106

Hing, N., Russell, A. M. T., Black, A., Rockloff, M., Browne, M., Rawat, V., … Woo, L. (2022). Gambling prevalence and gambling problems amongst land-based-only, online-only and mixed-mode gamblers in Australia: A national study. *Computers in Human Behavior*, *132*, 107269. https://doi.org/10.1016/j.chb.2022.107269

Hing, N., Smith, M., Rockloff, M., Thorne, H., Russell, A. M. T., Dowling, N. A., & Breen, H. (2022). How structural changes in online gambling are shaping the contemporary experiences and behaviours of online gamblers: An interview study. *BMC Public Health*, *22*(1), 1620. https://doi.org/10.1186/s12889-022-14019-6

Hing, N., Thorne, H., Russell, A. M. T., Newall, P. W. S., Lole, L., Rockloff, M., … Tulloch, C. (2024). 'Immediate access … everywhere you go': A grounded theory study of how smartphone betting can facilitate harmful sports betting behaviours amongst young adults. *International Journal of Mental Health and Addiction*, *22*(3), 1413–1432. https://doi.org/10.1007/s11469-022-00933-8

Hopfgartner, N., Auer, M., Griffiths, M. D., & Helic, D. (2022). Predicting self-exclusion among online gamblers: An empirical real-world study. *Journal of Gambling Studies*, *39*(1), 447–465. https://doi.org/10.1007/s10899-022-10149-z

Hopfgartner, N., Auer, M., Helic, D., & Griffiths, M. D. (2024). Using artificial intelligence algorithms to predict self-reported problem gambling among online casino gamblers from different countries using account-based player data. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-024-01312-1

Kairouz, S., Costes, J.-M., Murch, W. S., Doray-Demers, P., Carrier, C., & Eroukmanoff, V. (2023). Enabling new strategies to prevent problematic online gambling: A machine learning approach for identifying At-risk online gamblers in France. *International Gambling Studies*, *23*(3), 471–490. https://doi.org/10.1080/14459795.2022.2164042

Kock, F., Berbekova, A., & Assaf, A. G. (2021). Understanding and managing the threat of common method bias: Detection, prevention and control. *Tourism Management*, *86*, 104330.

Kuhn, M., & Wickham, H. (2020). *Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles*. [Computer software] https://www.tidymodels.org.

LaBrie, R., & Shaffer, H. J. (2011). Identifying behavioral markers of disordered internet sports gambling. *Addiction Research & Theory*, *19*(1), 56–65. https://doi.org/10.3109/16066359.2010.512106

Lakew, N., & Lindner, P. (2025). What do they say they are doing? A mixed-methods analysis of Swedish gambling operators' duty of care action plans. *Harm Reduction Journal*, *22*(1), 196. https://doi.org/10.1186/s12954-025-01349-y

Luquiens, A., Tanguy, M. L., Benyamina, A., Lagadec, M., Aubin, H. J., & Reynaud, M. (2016). Tracking online poker problem gamblers with player account-based gambling data only. *International Journal of Methods in Psychiatric Research*, *25*(4), 333–342. https://doi.org/10.1002/mpr.1510

Marionneau, V., Ristolainen, K., & Roukka, T. (2025). Duty of care, data science, and gambling harm: A scoping review of risk assessment models. *Computers in Human Behavior Reports*, *18*, 100644. https://doi.org/10.1016/j.chbr.2025.100644

Murch, W. S., Kairouz, S., Dauphinais, S., Picard, E., Costes, J.-M., & French, M. (2023). Using machine learning to retrospectively predict self-reported gambling problems in Quebec. *Addiction*, *118*(8), 1569–1578. https://doi.org/10.1111/add.16179

Percy, C., França, M., Dragičević, S., & d'Avila Garcez, A. (2016). Predicting online gambling self-exclusion: An analysis of the performance of supervised machine learning models. *International Gambling Studies*, *16*(2), 193–210. https://doi.org/10.1080/14459795.2016.1151913

Perrot, B., Hardouin, J. B., Thiabaud, E., Saillard, A., Grall-Bronnec, M., & Challet-Bouju, G. (2022). Development and validation of a prediction model for online gambling problems based on players' account data. *Journal of Behavioral Addictions*, *11*(3), 874–889. https://doi.org/10.1556/2006.2022.00063

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Sacco, P., & Jeong, J. (2025). Assessing the risk of problem gambling among lottery loyalty program members: A machine learning approach. *Addictive Behaviors*, *168*, 108372. https://doi.org/10.1016/j.addbeh.2025.108372

Swanton, T. B., Gainsbury, S. M., & Blaszczynski, A. (2019). The role of financial institutions in gambling. *International Gambling Studies*, *19*(3), 377–398. https://doi.org/10.1080/14459795.2019.1575450

The UK Behavioural Insights Team. (2021). *Gambling behaviour: What can bank transaction data tell us? The feasibility study.*

Part 2: Analysis of HSBC UK customer data. Behavioural Insights Team. https://www.bi.team/blogs/dealing-new-data-what-bank-transactions-can-tell-us-about-gambling-behaviour/.

UK Gambling Commission. (2022). Gambling commission sets new rules on action for at risk customers. Available at: https://www.gamblingcommission.gov.uk/news/article/gambling-commission-sets-new-rules-on-action-for-at-risk-customers.

UK Gambling Commission. (2023). Customer interaction guidance for remote gambling licensees (Formal guidance under SR code provision 3.4.3). https://www.gamblingcommission.gov.uk/licensees-and-businesses/guide/customer-interaction-formal-guidance-for-remote-gambling-operators.

Zendle, D., & Newall, P. (2024). The relationship between gambling behaviour and gambling-related harm: A data fusion approach using open banking data. *Addiction*, *119*(10), 1826–1835. https://doi.org/10.1111/add.16571