

[Request a Demo](#)

Prevent Churn in Online Sports

[GAMING](#) [MARKETING / SALES](#) [CHURN/RETENTION](#) [IMPROVE CUSTOMER EXPERIENCE](#) [INCREASE REVENUE](#)
[AUGMENTATION](#) [BINARY CLASSIFICATION](#) [CHURN / RETENTION](#) [END TO END](#)

Predict the likelihood of a player (i.e. individual placing a bet) going dormant in the course of 28 days.

[Build with Free Trial >](#)

[Overview](#) [Technical implementation](#) [Business Implementation](#)

Overview

Business Problem

The online gambling industry is one of the most revenue-generating branches of the entertainment business; in the US alone in 2019, it generated 40 billion dollars. Online betting platforms allow customers to bet on various games such as races (horse, greyhound, harness racing, etc.) and sports (American football, baseball, basketball, cricket, golf, etc.). Customer retention is a major issue within every online sports betting platform due to extreme competition. Since sports and games are unvarying across the various platforms, customer experience remains the most important factor to retention.

Online gaming services generally experience a high churn rate (around 40% churn) in the first week of their deposits or just after submitting a registration form, i.e., even before placing their first bet. These are the customers who only make one bet and never come back. Currently, marketing teams

approach these customers by looking at their betting amounts and win-loss ratio to determine whom to contact and what intervention strategy to apply. These interventions can vary from offering “deposit match” to giving “free bets.”

Intelligent Solution

AI can help predict the likelihood that a player will make at least 1 bet in the next 28 days. Models will be able to identify customers at risk so marketing teams can proactively intervene to influence customers’ behavior. Businesses would be able to reduce their player churn by getting a risk score for players and intervening those who have a high score. Through Prediction Explanations, businesses can understand the reasons behind those scores and then target the riskiest customers. Customer retention teams tailor interventions or offers using the information provided by these explanations. As an example, customer retention teams can offer 100% deposit match offers to customers who fall in the top 2 deciles, 50% to the mid deciles, and 25% to the bottom deciles.

Value Estimation

How would I measure ROI for my use case?

To calculate the ROI of this use case, we would need to benchmark the AI model results to your existing churn numbers. As an example:

- Let’s say current churn rate @ 30% = ~1000
- Reduction in churn rate = ~800
- Average earning from one bet = \$50 per week
- Cost of intervention = \$10
- Net profit $\$50 - \$10 = \$40$
- Weekly extra revenue generated $\$40 \times 200 = \$8,000$
- **Annually $(\$8,000 \times 52) = \$416,000$**

Technical Implementation

About the Data

For illustrative purposes, this tutorial uses a synthetic dataset that includes players’ past betting activities and demographic data. The features included were all synthetically developed.

Problem Framing

The target variable being predicted for this use case is to identify players who will go dormant in the next 28 days, meaning they will not make any bets throughout this duration. This choice in target makes this a binary classification problem.

The features we add to our model include data on the customer, past transactions, and platform activity. Beyond these features, we suggest incorporating any additional data your organization may collect that could be relevant to the use case. As you will see later, DataRobot is able to help you quickly differentiate important vs unimportant features.

Sample Feature List

Feature Name	Data Type	Description	Data Source	Example
Cust ID	Numeric	Customer Identification	Customer	False
Age	Numeric	Customer Age	Customer	42
Email	Text	Email Address	Email	xyz@gmail.com
Gender	Categorical	Gender	Customer	Male
Join Date	Date	Joining date of the player	Customer	28/01/2019
Deposit_date	Date	First Deposit Date	Date	29/01/2019
Day_Sign_up_flag	Binary	Sign_up and Deposit Day is same ,1 ,0	Transaction	No
First_Deposit_amount	Numerical	First Time Deposit Amount	Transaction	\$500
First Deposit Type	Categorical	First time bet type – Free Bet & Cash	Transaction	Cash
Racing_Bet	Numerical	Count of bets on racing (F1, Horse race)	Activity	3

Feature Name	Data Type	Description	Data Source	Example
Sports_Bet	Numerical	Count of bets on sports (Football, Cricket, Rugby)	Activity	2
Total_Bets	Numerical	Total bets placed	Activity	5
Tot_Sum_Bets	Numerical	Total Bets Amount	Activity	\$500
Max_Bet	Numerical	Maximum Bet Amount	Activity	\$200
Min_Bet	Numerical	Minimum Bet Amount	Activity	\$50
Total_Free_Bets	Numerical	Total Free Bets given to player	Activity	2
Tot_Free_bet_amt	Numerical	Total free bet amount in \$\$	Activity	\$20
Sum_Paid	Numerical	Amount paid to customer from winning bets	Activity	\$100
Days Since Last Bet	Numerical	Number of days since last placed bet	Activity	4
Total Weighted Average Price	Numerical	Weighted Avg Price of total amount played incl free bets	Activity	
Number of bets placed during Normal/Odd hours	Binary	Count of bets played 9am-9pm (Normal) 9pm-9am(Odd)	Activity	
Total_Withdrawn	Numerical	Total amt withdrawn	Activity	
Withdrawn_Hours	Numerical	Amount withdrawn between various hours (7am-5pm), (5pm-10pm),(10pm-7am)	Activity	

Feature Name	Data Type	Description	Data Source	Example
Win_Ratio	Numerical	Win/Loss	Activity	
Bets made yesterday	Categorical	Number of bets made t-1	Activity	1
Bets made 2 days ago	Categorical	Number of bets made t-2	Activity	1
Bets made 3 days ago	Categorical	Number of bets made t-3	Activity	1
Bets made 10 days ago	Numerical	Number of bets made in last 10 days	Activity	20
tenureDays	Numerical	Number of days from sign up	Customer	30
Category	Categorical	[Horse Race, F1, Soccer, Rugby]	Product	Soccer
Is_dormant (<i>TARGET</i>)	Categorical	If (player placed a bet within 28 days, 0,1)		No

Model Training

DataRobot Machine Learning automates many parts of the modeling pipeline. Instead of having to hand-code and manually test dozens of models to find the one that best fits your needs, DataRobot automatically runs dozens of models and finds the most accurate one for you, all in a matter of minutes. In addition to training the models, DataRobot automates other steps in the modeling process such as processing and partitioning the dataset.

For this use case, the dataset needed a partitioning strategy called *Group Partitioning* on the Cust ID.

Partitioning by Group ID ensures that all members of the group fall within the same partition. By grouping you learn only from all the observations in specific groups and predict on other groups. This allows you to better assess your performance on new customers that you have never seen before, and build models that are more robust to those new customers.

Advanced Options

Partitioning Smart Downsampling Time Series Feature Constraints Additional

Select partitioning method:

Random Partition Feature **Group** Date/Time Stratified

All rows with the same single value of the selected feature are in the same partition. Each partition can contain multiple values of the feature.

Group ID feature:

This must be a feature from the dataset.

...

Cust ID	Date	feat	Churn
123	1/1/2015		No
123	1/1/2016		No
123	1/1/2017		Yes
156	1/1/2015		No
156	1/2/1015		Yes
178	1/1/2015		No
178	1/2/2015		No
178	1/3/2015		No
178	1/4/2015		Yes

Cust ID	Date	feat	Churn
123	1/1/2015		No
123	1/2/2015		No
123	1/3/2015		Yes
156	1/1/2015		No
156	1/2/1015		Yes

Partition 1

Cust ID	Date	feat	Churn
178	1/1/2015		No
178	1/2/2015		No
178	1/3/2015		No
178	1/4/2015		Yes

Partition 2

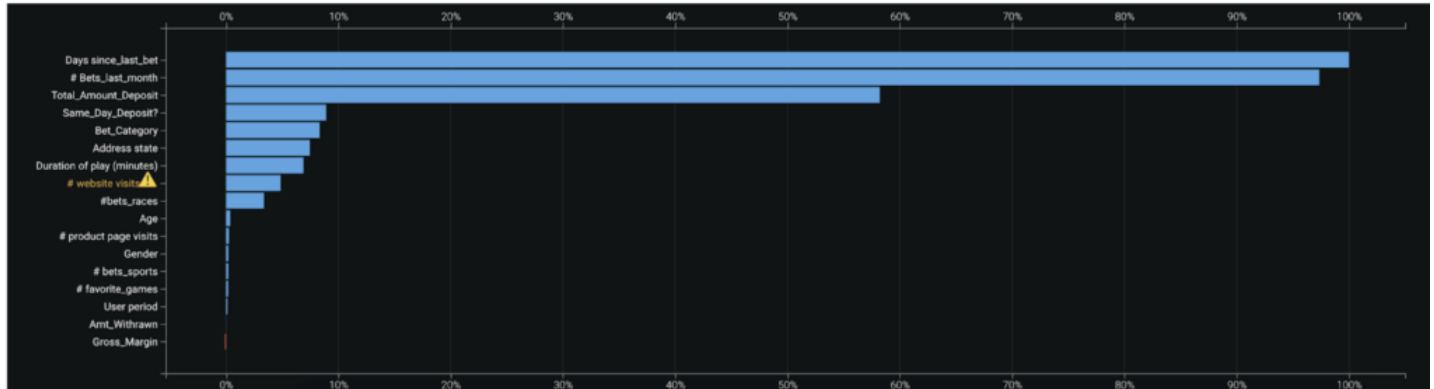
To understand why we need to partition the data, see [Training, Validation, and Holdout](#). Also, have a look at this [Churn Playbook](#) article for more information about Group Partitioning.

Interpret Results

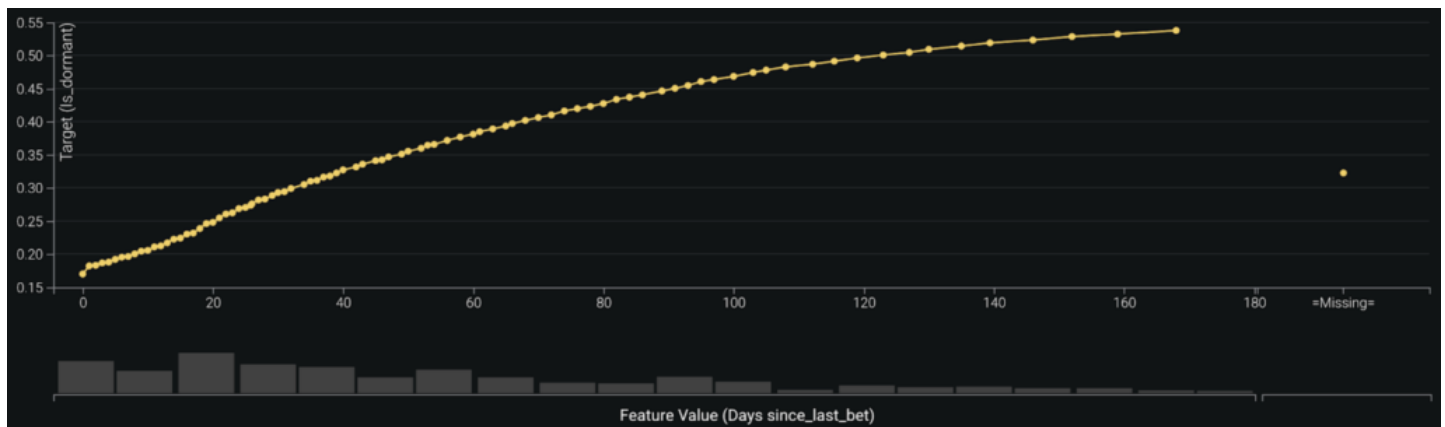
Feature Impact provides an understanding of feature importance. (You can read more in this [Feature Impact in Machine Learning](#) community article.)

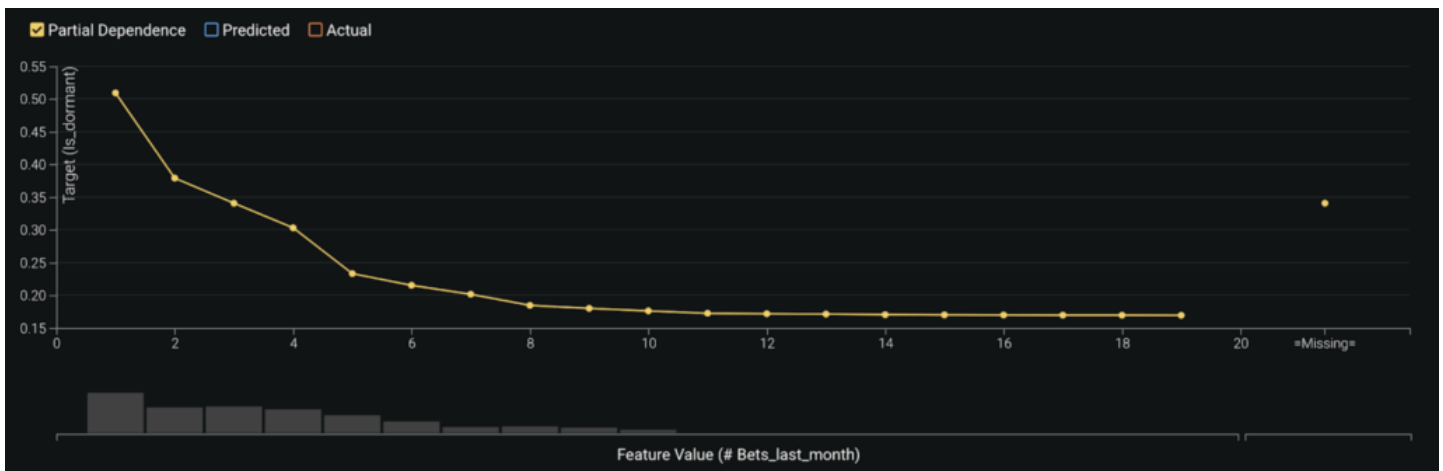
The magnitude of importance is ranked from most important feature on the top of the list to least important. In the chart below, **Days since last bet** is the most important feature, followed by **# Bets last month**, **Total Amount Deposit**, **Same Day Deposit**, etc.

Note : DataRobot has detected a redundant feature (**# website visits**) and flagged it for us with .



In assessing the partial dependence plots to further evaluate the marginal impact top features have on the predicted outcome, we learn that players who have placed their last bet recently are less likely to churn than players who have not placed a bet in a while. Also, the number of bets placed in recent months has an inverse relationship on the likelihood of churn. Additional insights can be discovered such as players who placed their bet on the day of signing up are less likely to churn and a few states are riskier than others.





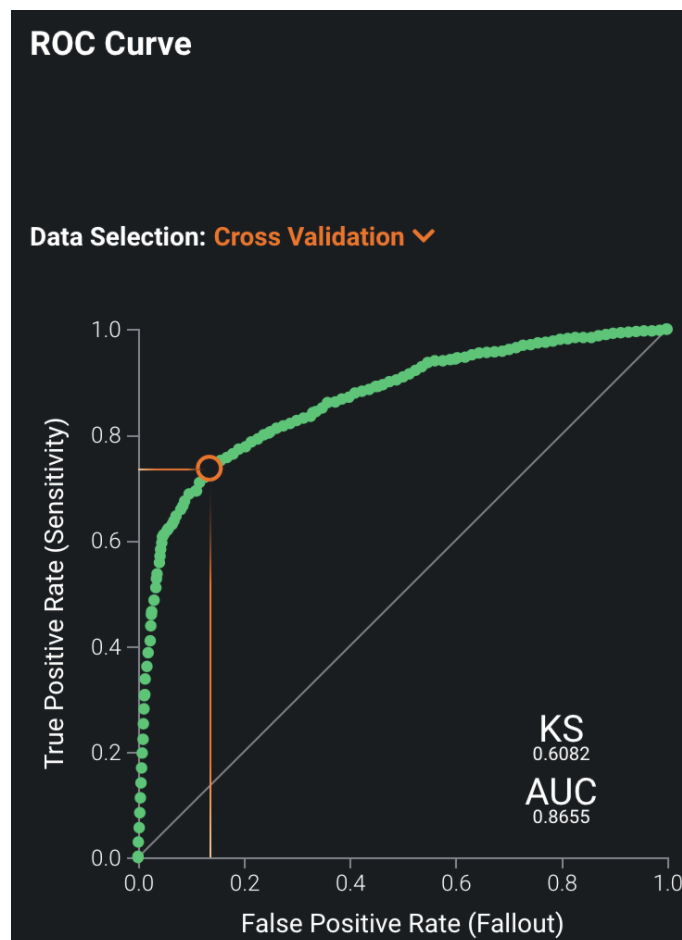
DataRobot's Prediction Explanations provide a more granular view to interpret the model results. (More information on Prediction Explanations is provided in [the public documentation](#).)

Here, we see why a given player was predicted to churn or not, based on the top predictive features.

ID	PREDICTION	EXPLANATIONS
2798	0.951	+++ Days since_last_bet = 172 +++ # Bets_last_month = 1 +++ Total_Amount_Deposit = 126
3055	0.944	+++ Days since_last_bet = 177 +++ # Bets_last_month = 1 +++ Total_Amount_Deposit = 405.5
711	0.944	+++ Days since_last_bet = 133 +++ # Bets_last_month = 1 +++ Total_Amount_Deposit = 452.5
2420	0.019	--- # Bets_last_month = 9 --- Days since_last_bet = 7 --- Same_Day_Deposit? = 'yes'
223	0.017	--- # Bets_last_month = 8 --- Days since_last_bet = 3 --- Total_Amount_Deposit = 1364
1581	0.016	--- Days since_last_bet = 6 --- # Bets_last_month = 7 --- Total_Amount_Deposit = 1200

Evaluate Accuracy

We want the model to learn the ranking of the probabilities so we can focus on the customers with high probability scores. Therefore, AUC was chosen as an optimization metric.



Churn use cases benefit from the model's ability to correctly predict as many True Positives as possible while also minimizing False Positives.

		Predicted		
		-	+	
Actual	-	1541 (TN)	240 (FP)	1781
	+	233 (FN)	652 (TP)	885
		1774	892	2666

Post-Processing

The probabilities given by the chosen model were exported to the data source which can then be provided to marketing teams. We used DataRobot's Lift Chart to identify which players to reach out to (those likely to churn) and which can be ignored (are most likely to convert).

Business Implementation

Decision Environment

After you choose the right model that best fits your data, DataRobot makes it easy to deploy the model into your desired decision environment. Decision environments are the methods by which predictions will ultimately be used for decision-making.

Decision Maturity

Automation | **Augmentation** | Blend

The model enables customer retention teams to focus their efforts on customers who can be retained by either giving them a call or sending them an email. This will make retention teams' jobs much easier and improve the overall conversion rate across the board.

Model Deployment

The model's output needs to be consumed in an actionable way to be able to get real value; otherwise it will turn into an experimental project with no tangible value to the business. The output of the model—a list of players which our model thinks are more likely to churn—will get sent to the **customer retention** team. The decision engine can either be a simple CSV file or integration with CRM systems; either way, DataRobot makes it easy for end users to use these predictions.

For instance, the predictions can be integrated with **Microsoft Power BI** to create a dashboard that can be accessed by the customer retention team to support decisions on prioritizing which customers to reach out to offer **free bets**. Models score the customer at risk overnight and send the resulting predictions to the Power BI dashboard. The list of those customers, their propensity score, and Prediction Explanations associated with the score will be sent to the retention call center.

Decision Stakeholders

Decision Executors

Decision executors consume the predictions and make decisions on a daily/weekly basis. This can be a member of:

- Retention Team
- Marketing Team
- Customer Service Team

Decision Managers

Decision managers are the executive stakeholders who will monitor and manage the program to analyze the performance of the rate of customer churn.

- Chief Marketing Officer
- Customer Experience Officer
- Manager Customer Engagement

Decision Authors

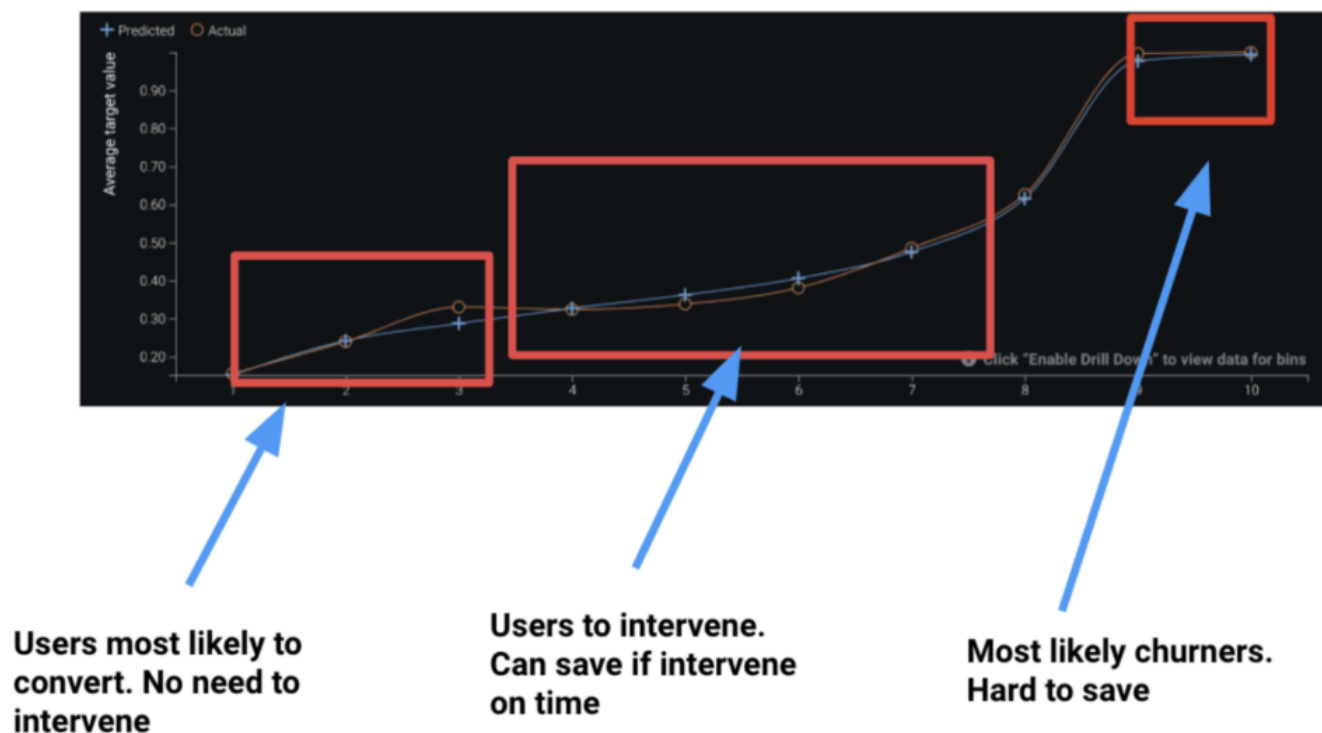
Decision authors are the technical stakeholders who will set up the decision flow in place.

- Data Scientist
- Customer UX analyst
- Customer engagement analyst
- Marketing Analyst

Decision Process

Different thresholds can be set to decide which intervention strategy to implement. These intervention strategies may include: sending a notification about an upcoming race or game, giving a player a “free bet,” offering a “deposit match offer,” calling the player to give a customized offer, etc.

These thresholds depend on companies’ risk appetites and the profitability of the books.



Assigning a different intervention to each cohort of players can be beneficial and reduces the unnecessary expenditure on those who are to influence.

High Risk: These players have a high likelihood of churn and likely have already made up their minds to leave. The retention team will have to work very hard to save these players.

Medium Risk: These are the cohort of players who can be influenced; the retention team should focus on this group. These players have stopped betting because they are looking for competitive offers; therefore, once the retention team gives them an extra “free bet” or DMO, these players will place the bets again.

Low Risk: These players can be saved by sending touch base emails or giving them one-off free bets. Intervention costs are low and conversion rate is high.

Model Monitoring

Decision Operators: IT/System Operations, Data Scientists

Prediction Cadence: Batch predictions generated on a daily basis

Model Retraining Cadence: Models retrained once data drift reaches an assigned threshold. Otherwise, retrain the models at the beginning of every new operating quarter.

Implementation Risks

Unsuccessful Intervention Strategies—Player decisions to stay or leave depend on the interventions decided by the marketing managers. If these interventions are not properly designed and implemented, then even if the model is highly accurate, the business ROI would still be low.

Trusted AI

In addition to traditional risk analysis, the following elements of AI Trust may require attention in this use case.

Target leakage: Target leakage describes information that should not be available at the time of prediction being used to train the model. That is, particular features may leak information about the eventual outcome that will artificially inflate the performance of the model in training. This use case requires the aggregation of historical data, making it vulnerable to potential target leakage. In the design of this model and the preparation of data, it is pivotal to identify the point of prediction and ensure no data be included past that time. DataRobot additionally supports robust target leakage detection in the second round of exploratory data analysis and the selection of the Informative Features feature list during autopilot. (Learn more [here](#) about target leakage.)

Experience the DataRobot AI Platform

Less Friction, More AI. Get Started Today With a Free 30-Day Trial.



[Sign Up for Free >](#)

Explore More Gaming Use Cases

The gaming industry leverages AI to predict user behavior, optimize odds, and ensure fair play. From enhancing online sports betting experiences to detecting fraudulent activities, AI plays a pivotal role in ensuring a secure and engaging environment for players.

[Gaming Use Cases >](#)

Explore More Use Cases

[GAMING](#)

Predict Player Lifetime Value ...

Predicting the LTV of
an individual player
for a casino / sports-
book.

[Learn More >](#)

[GAMING](#)

Predict Harmful Gambling...

Identify the
customers most likely
to exhibit harmful
gambling behavior.

[Learn More >](#)

[GAMING](#)

Fraud Detection in Online Sports

Predicting fraudulent
activity on an online
casino / sportsbook
platform.

[Learn More >](#)

Who Is DataRobot?

DataRobot is the leader in Value-Driven AI – a unique and collaborative approach to AI that combines our open AI platform, deep AI expertise and broad use-case implementation to improve how customers run, grow and optimize their business.



Gartner® Peer Insights™ 'Voice of the Customer': Data Science and Machine Learning Platforms, Peer Contributors, 20 Aug 2021

GARTNER and The GARTNER PEER INSIGHTS CUSTOMERS' CHOICE badge is a trademark and service mark of Gartner, Inc. and/or its affiliates and is used herein with permission. All rights reserved. Gartner Peer Insights Customers' Choice constitute the subjective opinions of individual end-user reviews, ratings, and data applied against a documented methodology; they neither represent the views of, nor constitute an endorsement by, Gartner or its affiliates.

© 2023 DataRobot, Inc.

[Legal](#)

[Privacy](#)

[Trust](#)