Peeking into the black box: An introduction to explainable AI (XAI)

10 December 2025

differential labs.

differential labs.

Mana Azizsoltani
Head of Data Science

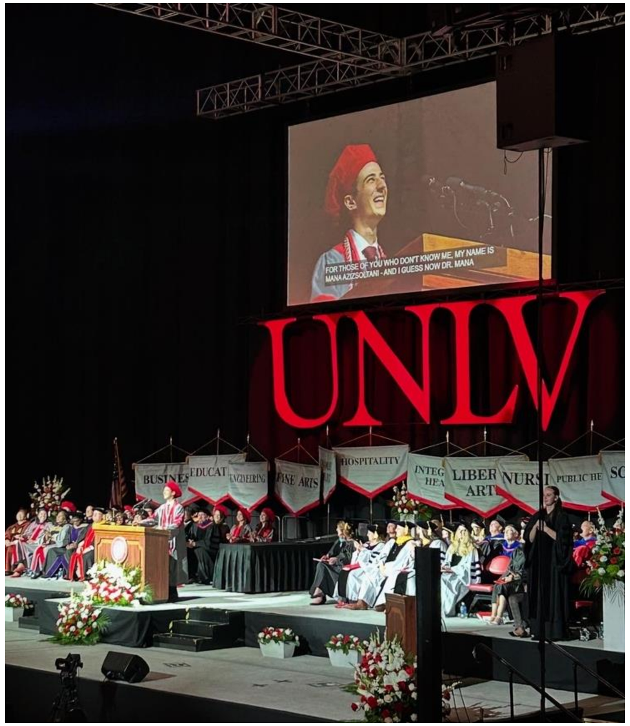
mana@diffgaming.com
+1 (702) 445-3737
https://www.linkedin.com/in/manaazizsoltani/

PRE-PREFACE

A little bit about myself

BACKGROUND

I am a trained data scientist and researcher-turned-consultant



I finished a BS in mathematics and MS in statistics to become a data scientist…

… to later complete a PhD in hospitality and be a university professor…

… to become a consultant and help casinos far and wide make more money with AI

I have done some of the publishing thing, and have some more papers loading…

Mana Azizsoltani
Differential Labs
Verified email at diffgaming.com
machine learning gambling revenue management casino

TITLE

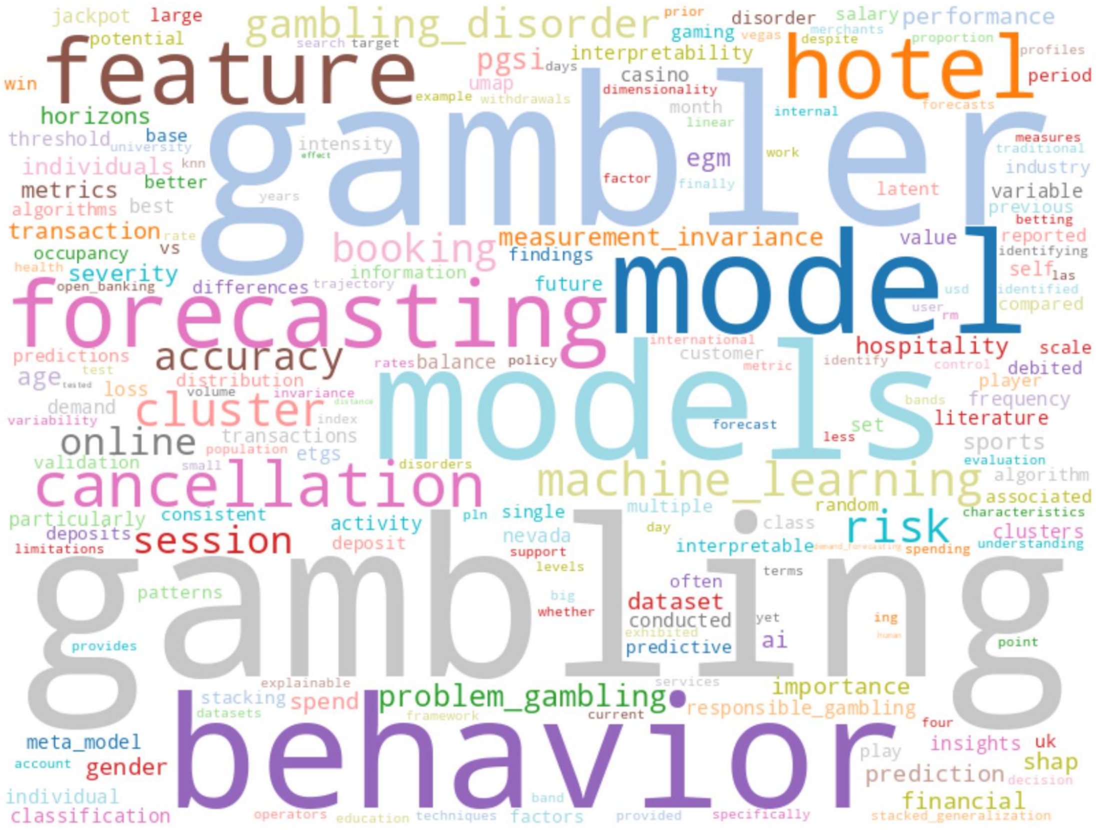
Towards Explainable Artificial Intelligence in Machine Learning: A study on efficient Perturbation-Based Explanations
I Gómez-Talal, M Azizsoltani, L Bote-Curiel, JL Rojo-Álvarez, A Singh
Engineering Applications of Artificial Intelligence 155, 110664

Across the Bettor-Verse: an Open Banking Perspective on Gambling in the United Kingdom
K Ghaharian, J Peterson, M Azizsoltani, RJ Young, ER Louderback
Journal of Gambling Studies, 1-18

Measurement invariance of the Problem Gambling Severity Index across sociodemographics and gambling modalities
M Azizsoltani, K Ghaharian, S Kraus, J Grubbs
International Gambling Studies, 1-19

Machine Learning in Hospitality: Interpretable Forecasting of Booking Cancellations
I Gómez-Talal, M Azizsoltani, P Talón-Ballestero, A Singh
IEEE Access

Payments Data in Gambling Research
K Ghaharian, M Azizsoltani



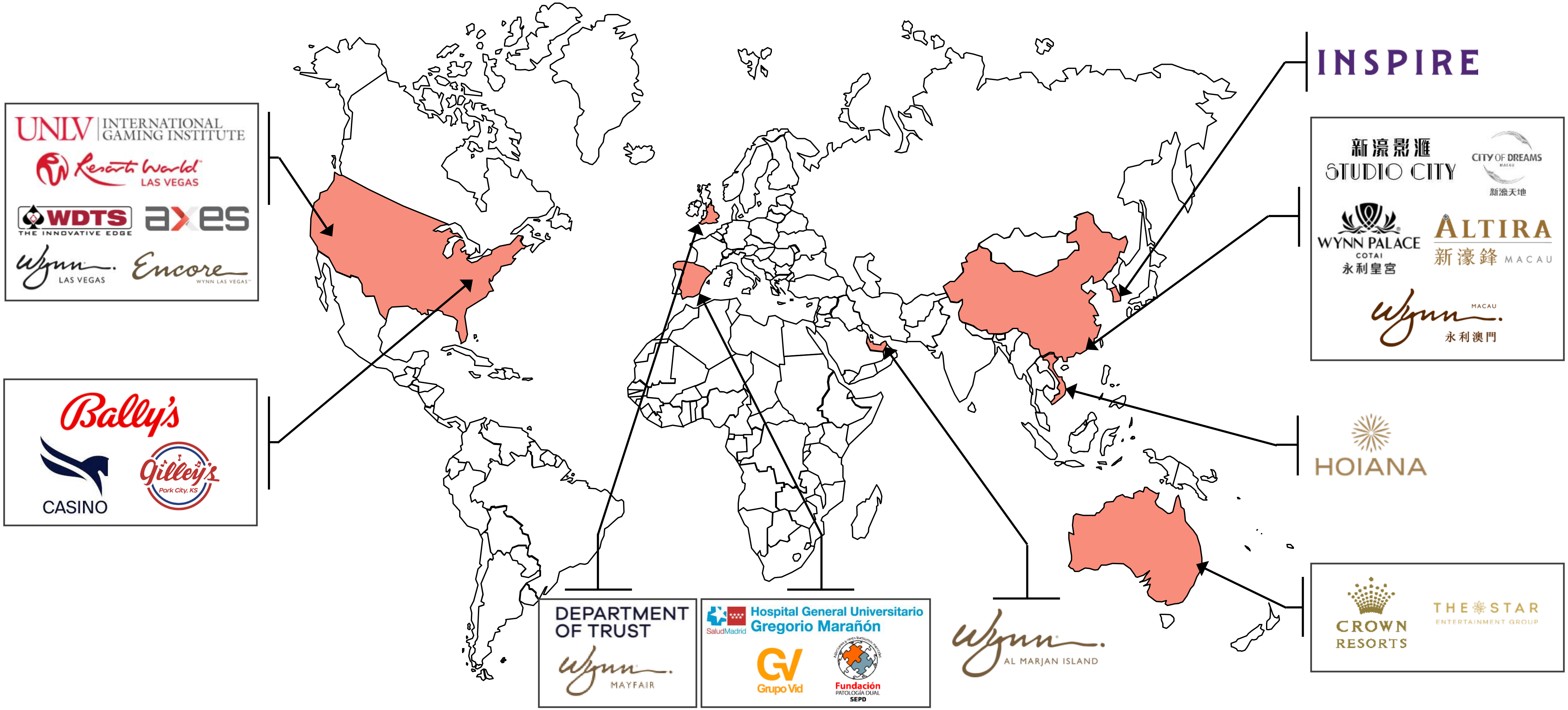
I also very much like to work with students and young professionals, serving in a variety of capacities

… But nowadays I am focused on the intersection between analytics/AI and the different facets of gambling

Casino Profitability			Public Health	
<u>Marketing</u>	<u>Surveillance</u>	<u>Revenue Optimization</u>	<u>Responsible Gaming</u>	<u>Psychiatry</u>
• Prediction of players likely to churn • Promotional preference hyper-personalization • Lifecycle modeling of database players • LLM offer generation • Detection of high value players • Determine players' sensitivity to offers	• Detection of suspicious behavior • Identification of promotional abusers • Players betting against the trends • Agile detection of advantage players • Determining the count(ability) of a particular game	• Dynamic pricing of table minimums • Hotel rooms pricing • Upselling and inventory management • Discount/rebate on loss programs • Casino floor optimization	• Detecting high risk gamblers based on behavior • Determining affordability of a gambler • Profiling of high-risk patrons	• Classifying gambling disorder patients based on their clinical data • Profiling dual gambling disorder patients based • Uncovering genetic markers of gambling disorder patients

EXPERIENCE

I have worked on international projects spanning various industries



PREFACE

A little bit about yourselves

Do not edit
How to change the design

Join at slido.com
#34426983

The Slido app must be installed on every computer you're presenting from

slido

ABOUT YOU

Tell me a bit about yourselves…

Do not edit
How to change the design

What is your area of expertise?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design

What models do you use or are you familiar with?

Section 1	A primer on XAI
Section 2	Industry examples of XAI
Section 3	Academic applications & demonstration

SECTION ONE

A primer on explainable AI

We will begin by talking about AI and then dive deeper

Artificial Intelligence

- What is AI?
- What are the different types?

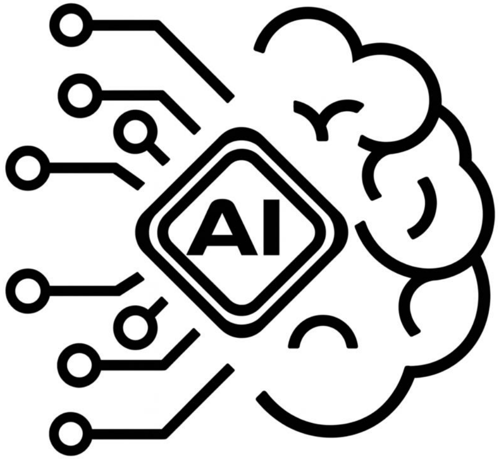
Explainable AI (XAI)

- What is XAI?
- Why do we need it?

Applications of XAI

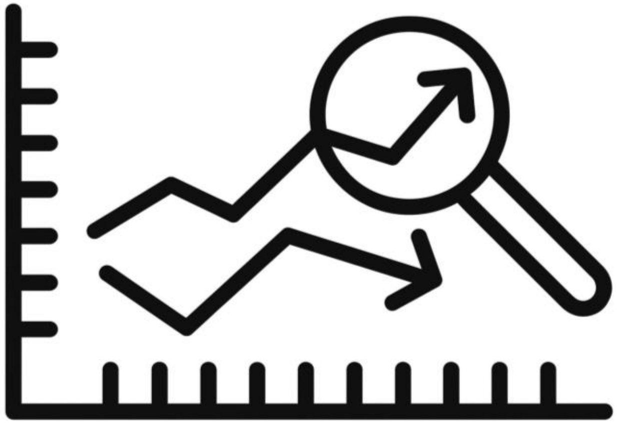
- How can XAI help us?
- Novel applications from work
- Academic applications

There are broadly two types of AI



Generative AI

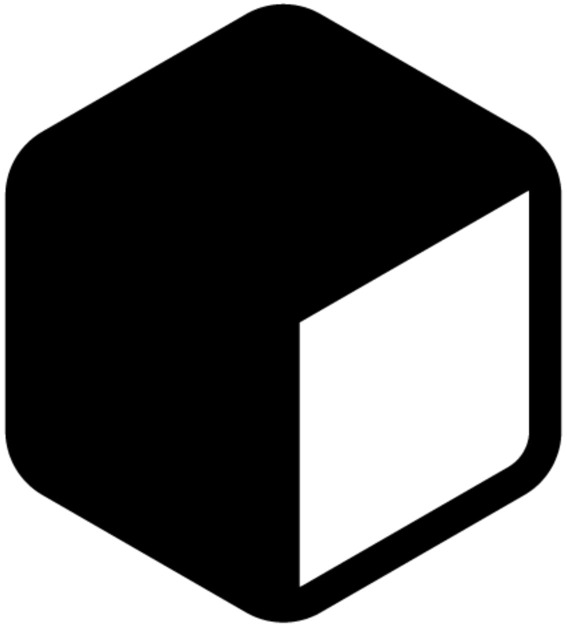
Creates new content from existing data, imitating human creativity.

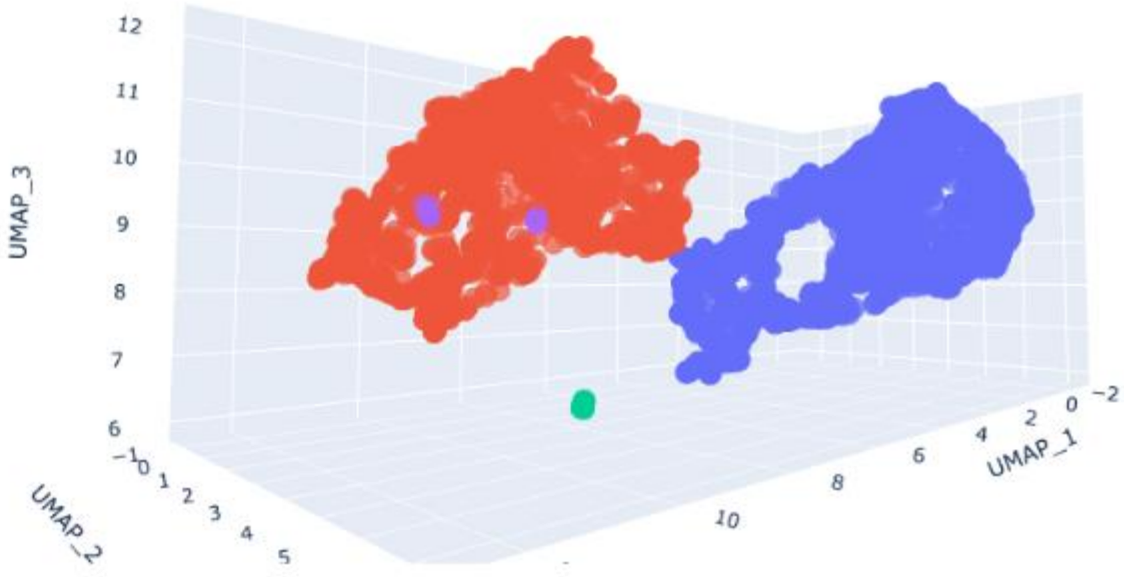


Predictive AI

Predicts future outcomes, for better decision making

But most AI algorithms are black-box models





"black box": we cannot see inside the AI models. How do models determine predictions?

We see clusters, but how do we know why a player is purple, for example?

To demonstrate: Can we predict the prestige of a professor?

This is our input dataset… three X variables, one Y variable

job	type	income	education	prestige
accountant	prof	62	86	82
pilot	prof	72	76	83
architect	prof	75	92	90
author	prof	55	90	76
chemist	prof	64	86	90
minister	prof	21	84	87
professor	prof	64	93	93
dentist	prof	80	100	90
reporter	wc	67	87	52

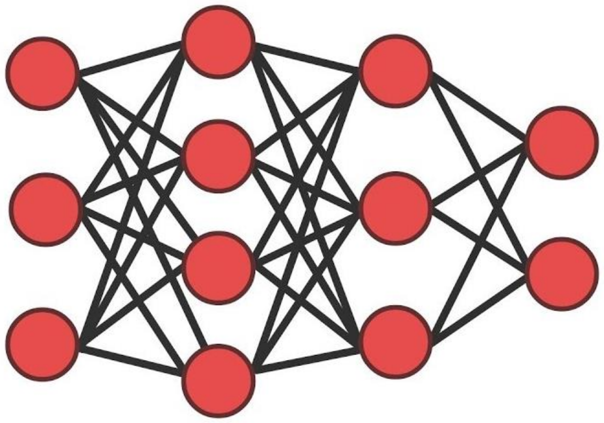
X vars

Target

OLS Regression

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.18503    3.71377  -0.050  0.96051
education     0.34532    0.11361   3.040  0.00416 **
income        0.59755    0.08936   6.687 5.12e-08 ***
typeprof     16.65751    6.99301   2.382  0.02206 *
typewc      -14.66113    6.10877  -2.400  0.02114 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
```

Neural Network



Prestige Prediction

97.5

93.4

But why do we care about a black box if it gives good results??

Ethics

amazon Apple

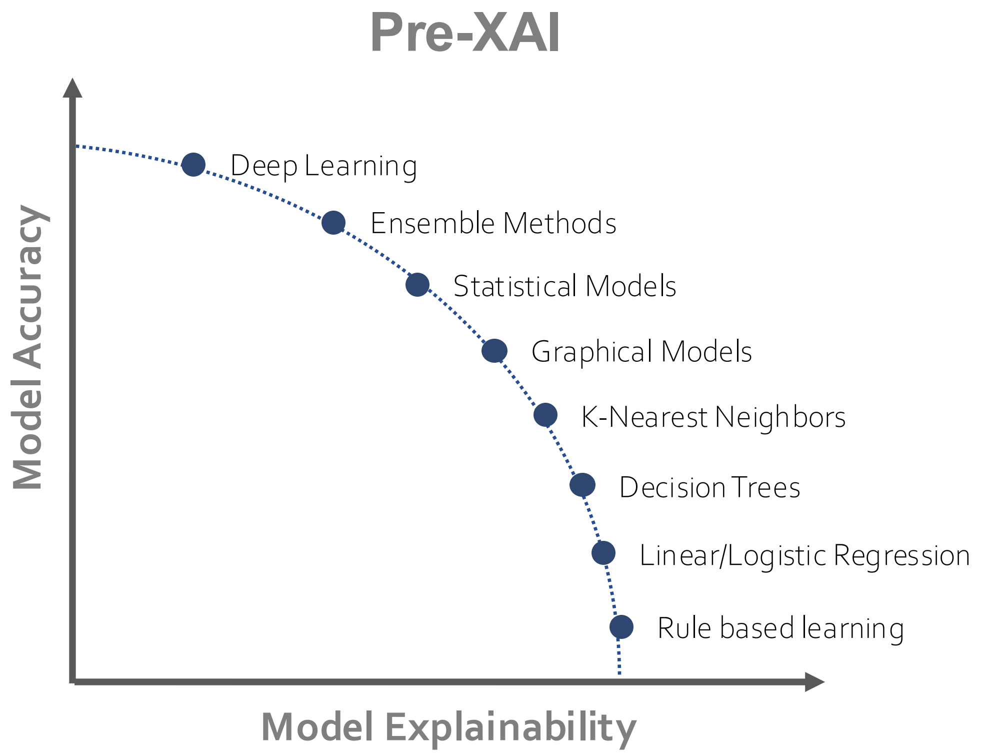
Regulation

EU Artificial Intelligence Act

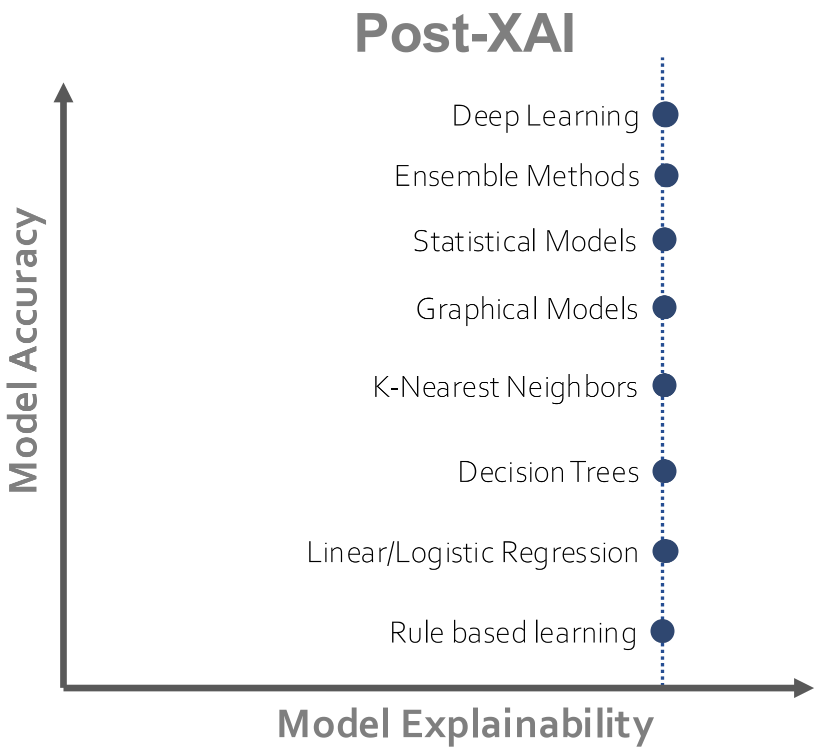
Auditability

Strategy

Traditional thinking suggested explainability came at the price of accuracy, but this is not the case anymore…

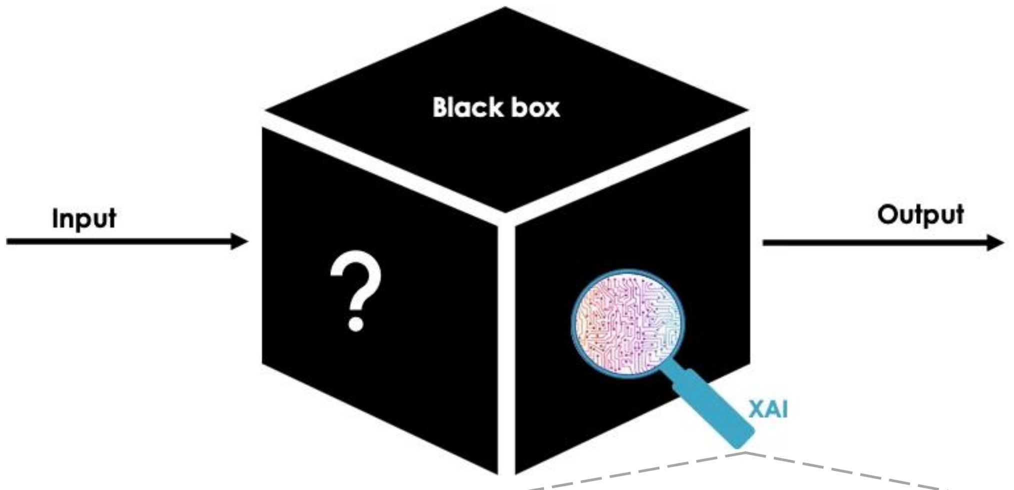


To get better predictions, we had to use deeper, more complex, and less interpretable models…



… But now, XAI has levelled the playing field on the Explainability dimension

Explainability in AI allows us to understand why a model gives a particular prediction



Explainability allows us to extract the "why" behind the prediction

A Unified Approach to Interpreting Model Predictions

Scott M. Lundberg
Paul G. Allen School of Computer Science
University of Washington
Seattle, WA 98105
slund1@cs.washington.edu

Su-In Lee
Paul G. Allen School of Computer Science
Department of Genome Sciences
University of Washington
Seattle, WA 98105
suinlee@cs.washington.edu

Contents lists available at ScienceDirect
Engineering Applications of Artificial Intelligence
journal homepage: www.elsevier.com/locate/engappai
ELSEVIER

Towards Explainable Artificial Intelligence in Machine Learning: A study on efficient Perturbation-Based Explanations

Ismael Gómez-Talal [a,b,*], Mana Azizsoltani [d,e,f], Luis Bote-Curiel [a], José Luis Rojo-Álvarez [a,c], Ashok Singh [d]

"Why Should I Trust You?" Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro
University of Washington
Seattle, WA 98105, USA
marcotcr@cs.uw.edu

Sameer Singh
University of Washington
Seattle, WA 98105, USA
sameer@cs.uw.edu

Carlos Guestrin
University of Washington
Seattle, WA 98105, USA
guestrin@cs.uw.edu

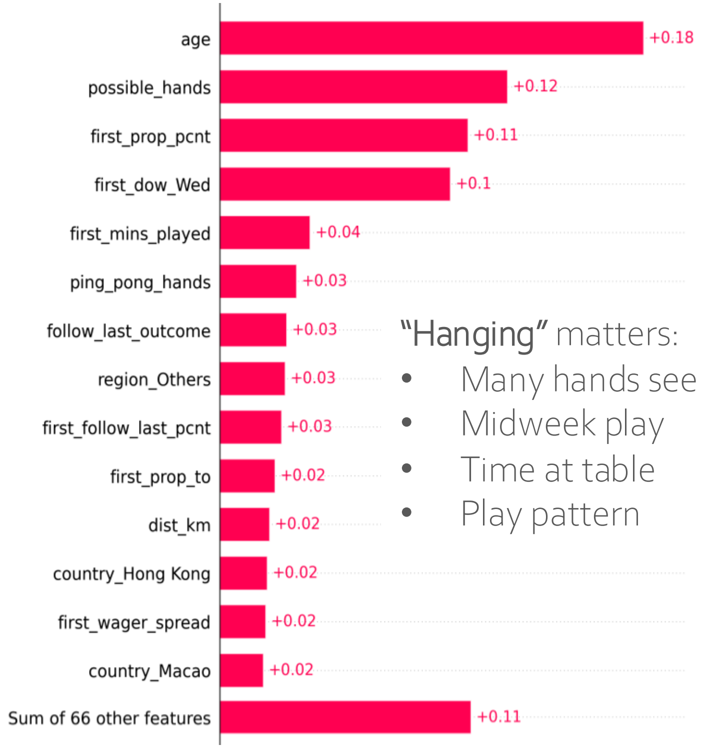
Explainability comes in many different shapes and sizes. We like to remember SMUD

Domain	Definition	Types
Scope	Describes how much of the model we're trying to understand. • *Do we want to explain the model's logic overall, or just a specific prediction?*	• Global: how does the entire model behave? • Local: how did the model make a particular prediction?
Mechanism	Describes when the explainability is achieved. • *Does the explanation happen within the model itself, or do we explain it after training?*	• Intrinsic: model is explainable by design • Post-hoc: explainability is applied after training
Understandability	Describes who can understand the model, • *Is it explainable to ML engineers or for the average Joe?*	• Explainable: someone knows what the model is doing • Interpretable: we can know what the model is doing
Dependence	Describes for which model the explainability method works. • *Does the method work for any model or only certain types?*	• Model specific: works only with a specific model or model family • Model agnostic: works for any model

GLOBAL VS. LOCAL

Global yields overall drivers, but local explanations unlock key insights. For example, high frequency prediction:

Global Explainability
(all players)

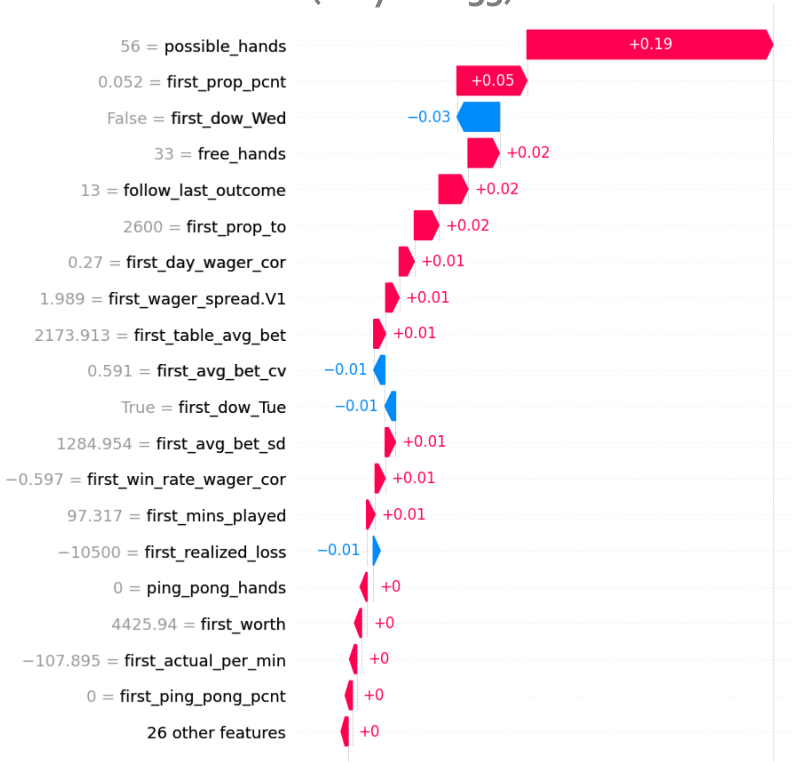


"Hanging" matters:

- Many hands seen
- Midweek play
- Time at table
- Play pattern

*"these variables influence predictions across **all players**"*

Local Explainability
(Player 1235)



*"We believe **player 1235** is a high frequency player because..."*

Some explainability is built into the model by definition. Back to our prestige example…

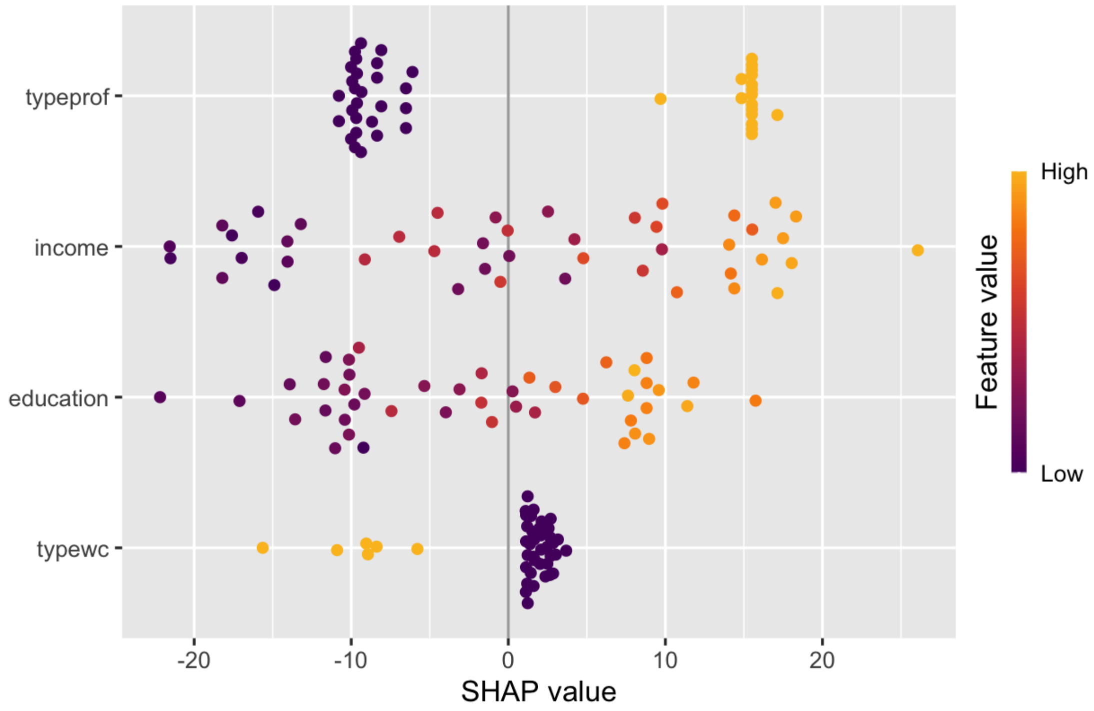
Intrinsic / specific
(linear regression)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.18503    3.71377  -0.050  0.96051
education    0.34532    0.11361   3.040  0.00416 **
income       0.59755    0.08936   6.687 5.12e-08 ***
typeprof    16.65751    6.99301   2.382  0.02206 *
typewc     -14.66113    6.10877  -2.400  0.02114 *
```

We can see the effect of a particular variable on the prediction

And some sort of variable importance measure

Post-Hoc / agnostic
(SHAP)



! This can be done with ANY model

Explainability is generalized insights, but interpretability is the human understandability of those insights

Explainability

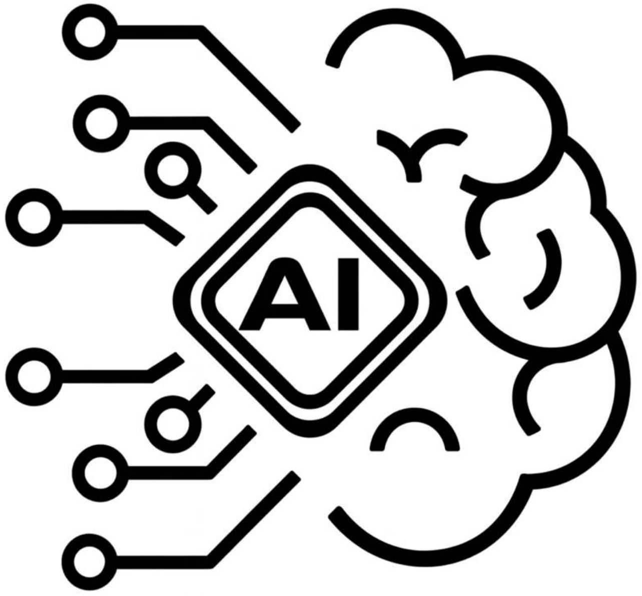
it's one thing for the analyst to be able to understand it…

Interpretability

… But it's different for the executives, operations, sales, marketing, IT teams to understand and make actionable

So, what about my bestie ChatGPT???

"So where does generative AI fit into the XAI equation?"



- What kind of model is it?
- Is it interpretable? Explainable?
- If so, how? Locally? Globally?
- Why would I talk about it in the context of XAI?

SECTION TWO

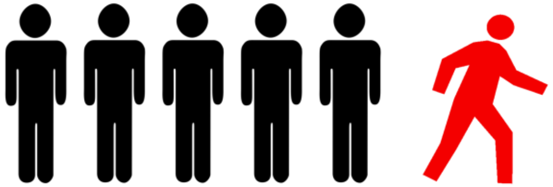
Explainable AI in the industry

I have selected a couple of case studies across a couple of broadly generalizable domains of gambling

Responsible Gambling

Prediction of problematic gambling behaviors

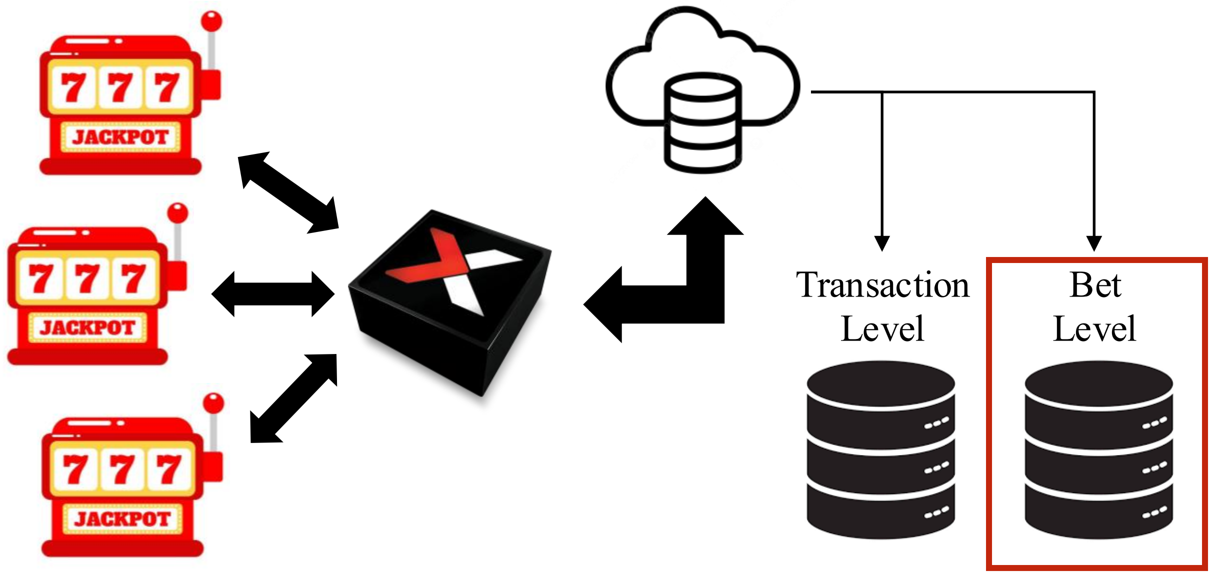
Churn Modeling



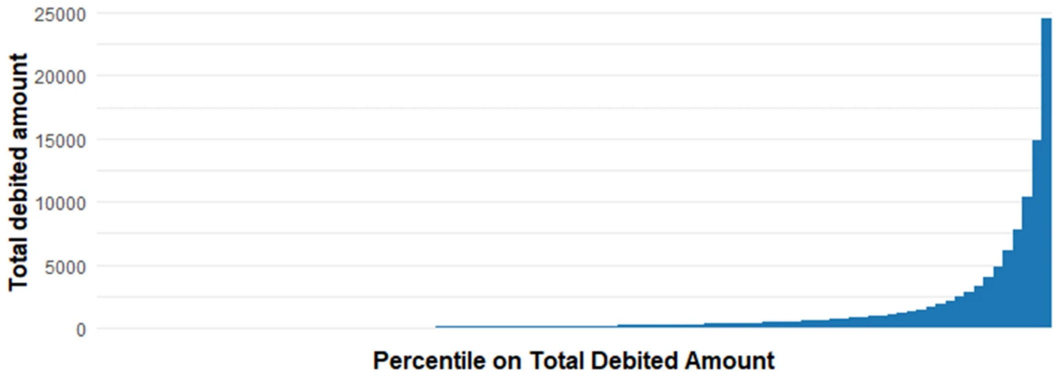
Is the player not going to return?

Many times, we don't have clinically-validated scales, so we try to identify problematic gamblers from their play behavior…

We collected bet-by-bet level data from 200k slot machine players in Europe…



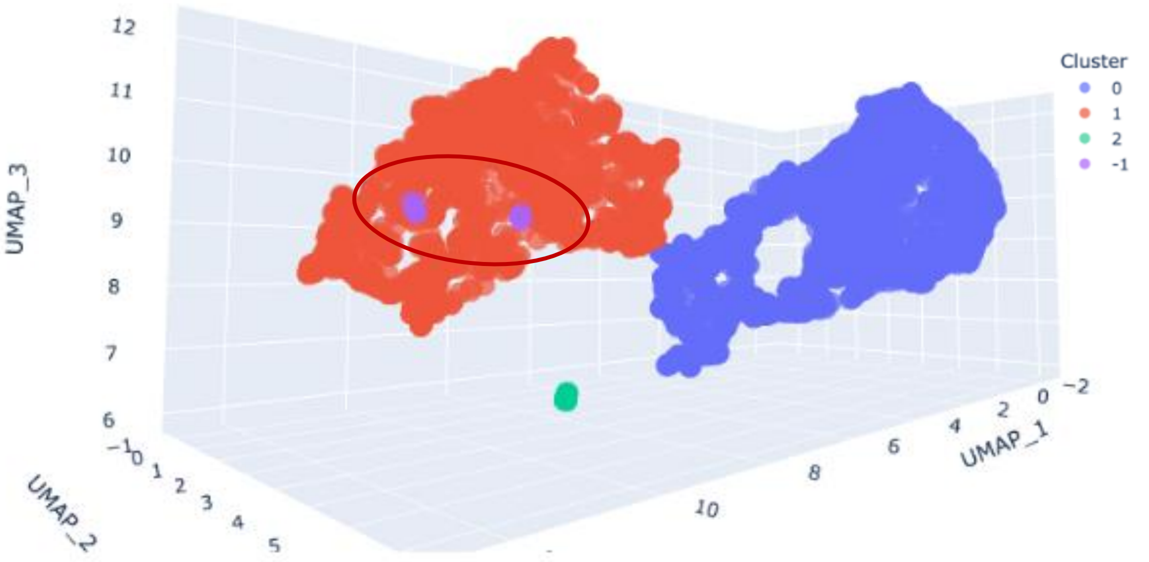
… and we crunched it to see if we could pick out the "highly involved" players



Gambling databases are highly skewed so we usually pick the "disproportionate" top %

Using explainable AI, we were able to unpack the clusters, looking at why clusters were formed

The difference in proximity of the purple outlier groups to the main group could hint that some particularly dangerous play could spawn from more "typical" highly involved activity



Feature Importance by Cluster

This was further seen when we broke down the clusters

Cluster -1: Outliers		Cluster 0: High Frequency		Cluster 1: High Spend		Cluster 2: Outliers	
Feature	**Imp.**	**Feature**	**Imp.**	**Feature**	**Imp.**	**Feature**	**Imp.**
Avg # of withdrawals/session	0.0643	Max # of monthly sessions	0.1242	Max # of monthly sessions	0.1177	Avg time btwn transactions	0.2553
Night time session percent	0.058	Avg monthly # of sessions	0.1072	Avg monthly # of sessions	0.1081	Avg time btwn deposits	0.2149
Avg balance trajectory	0.0576	Total # of sessions	0.0886	Avg net win/loss per session	0.0908	Std dev of time btwn deposits	0.044
Avg time btwn deposits	0.0542	Avg net win/loss per session	0.0886	Total # of sessions	0.0877	Avg elapsed hrs btwn sessions	0.0407
Std dev of time btwn deposits	0.054	Total net win/loss	0.074	Total net win/loss	0.0736	Std dev of time btwn sessions	0.035
Avg elapsed hours btwn sessions	0.0476	Avg # of withdrawals/session	0.0822	Avg # of withdrawals/session	0.0809	Avg monthly # of sessions	0.0336
Max win/loss per session	0.0472	Avg balance trajectory	0.0807	Avg balance trajectory	0.0715	Total # of sessions	0.0315
Avg net win/loss per session	0.0469	Avg time btwn deposits	0.0736	max win/loss per session	0.0701	Avg net win/loss per session	0.0304
Avg monthly # of sessions	0.0463	Avg elapsed hrs btwn sessions	0.0719	Avg elapsed hrs btwn sessions	0.0639	Total net win/loss	0.0289
Avg weekly # of sessions	0.0431	Night time session percent	0.0651	Avg time btwn deposits	0.063	max win/loss per session	0.0271

And then we used an LLM to help us inform the stakeholders automatically

We trained the LLM to think like an RG professional…

… and how to write an email

Mana Azizsoltani 9:10 PM
RG Alert: Patron 1235
To: Mana Azizsoltani

Hello,

Our behavioral risk monitoring system has flagged **Patron 1235** with elevated indicators consistent with potential gambling-related harm. This player's last session was at Local 75, and they frequently play at Local 124.

Reasons we are concerned about potential risk:

- Increased session length and frequency relative to baseline
- Volatility in recent play paired with loss-chasing behavior
- Moderate increase in night sessions
- Volatility in time between trips

Reasons this player may not be at elevated risk:

- Losses are not substantially out of ordinary.
- No observed changes in preferred games or betting style
- Player had a large win ($500)

Please limit direct contact or promotional offers to this player until further notice, and discreetly monitor their activity for any signs of distress or erratic behavior. This player's card has been temporarily deactivated for a cool-off period, and no player reinvestment, comp issuance, or direct marketing should be initiated until reactivation is confirmed.

Thank you for your attention to this matter.

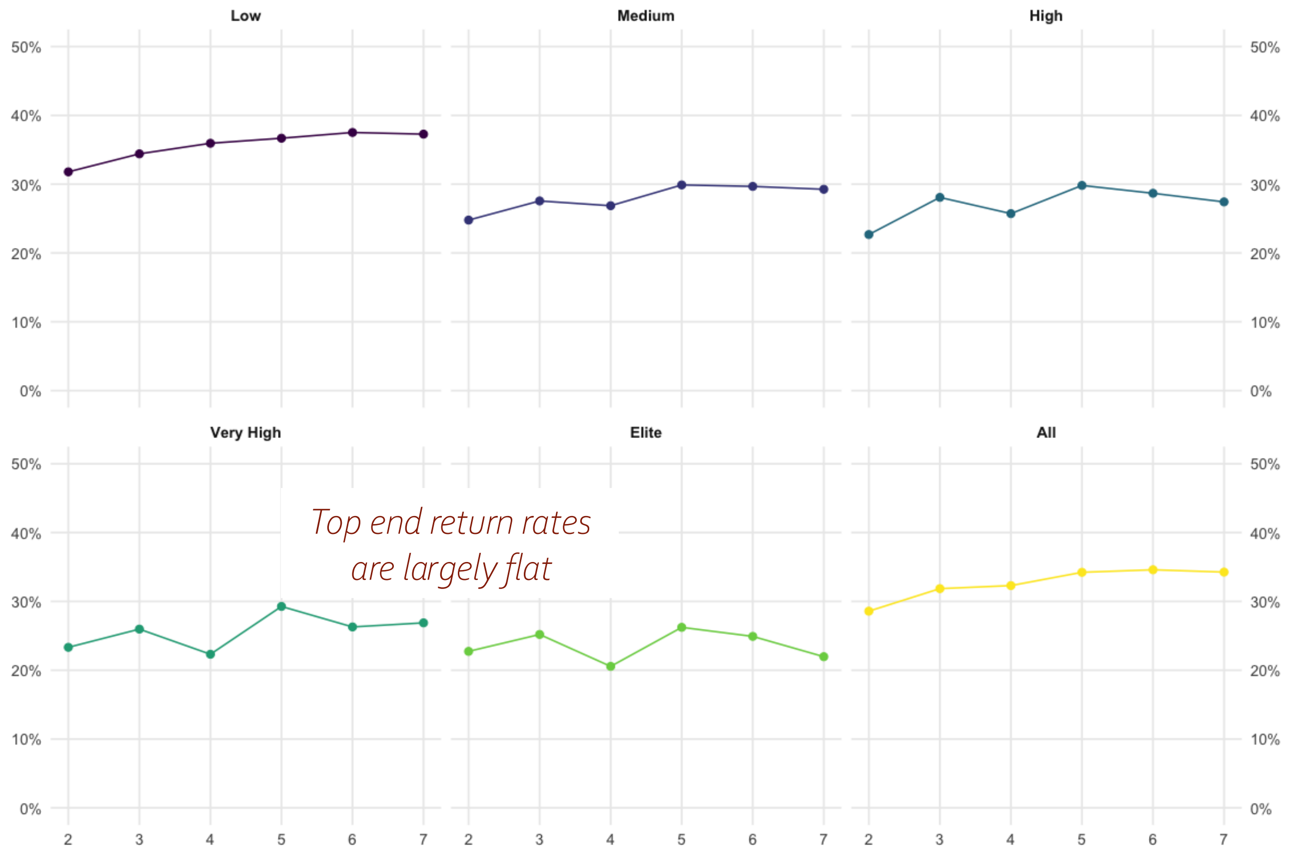
R(obot)G

1. Patron is flagged and report comes on latest data
2. Reasons why the model believes this prediction to be strong (positive shap) and rationale for scepticism (negative shap)
3. Dissemination of information and update on player status

The company in question had a small, gradual increase in churn, putting it at about 35%

Percent not returned next three-month period by ADT bin

% of patrons from previous qtr returning in the subsequent qtr



3-month period

Lowest end patrons have higher non-return rates

differential labs

We extracted the most important variables for churn prediction using explainable AI

Most important variable is regularity of visits

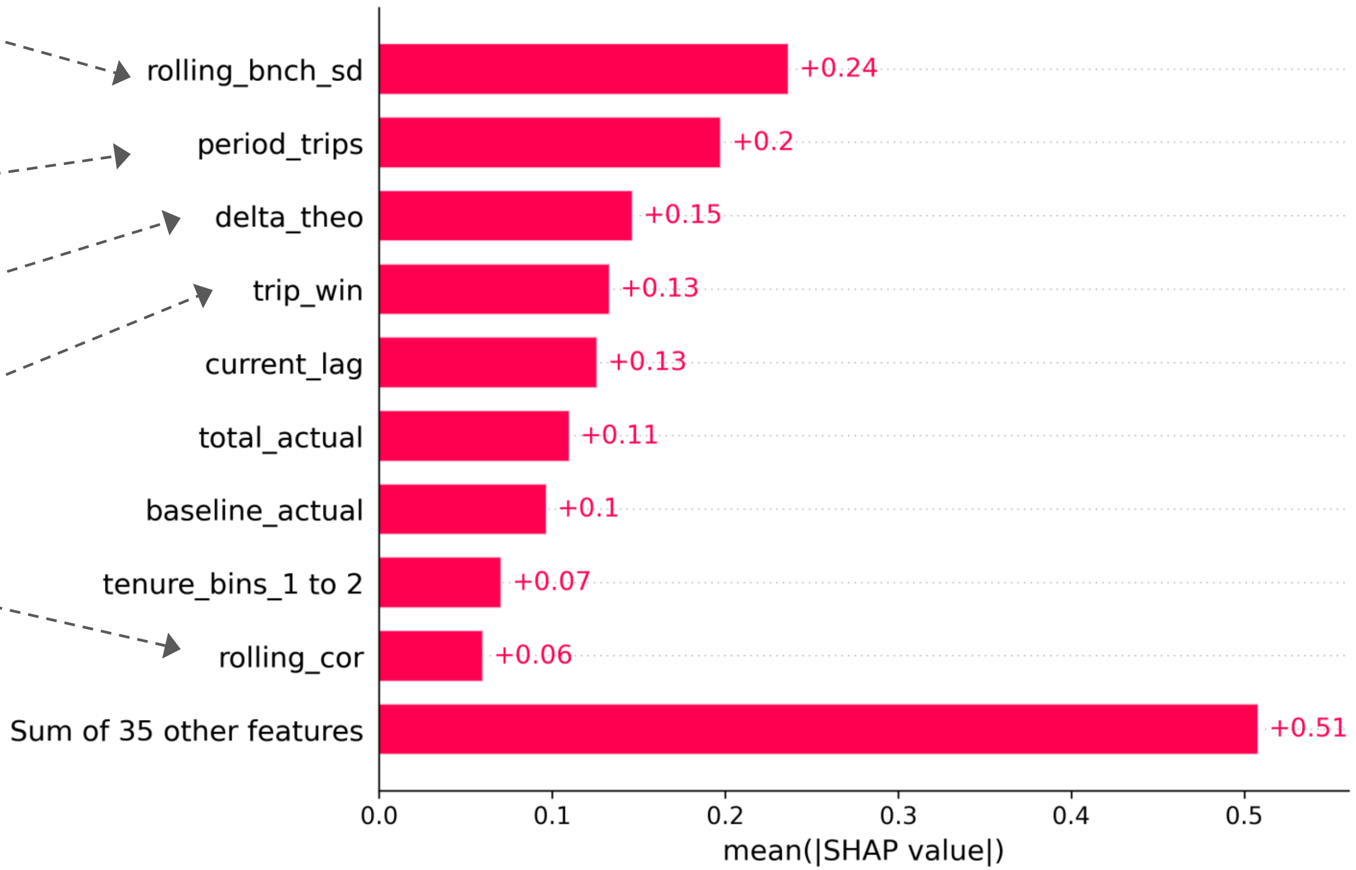
Recency is strong predictor

Change in theo win

Recent wins and losses

Frequency trend

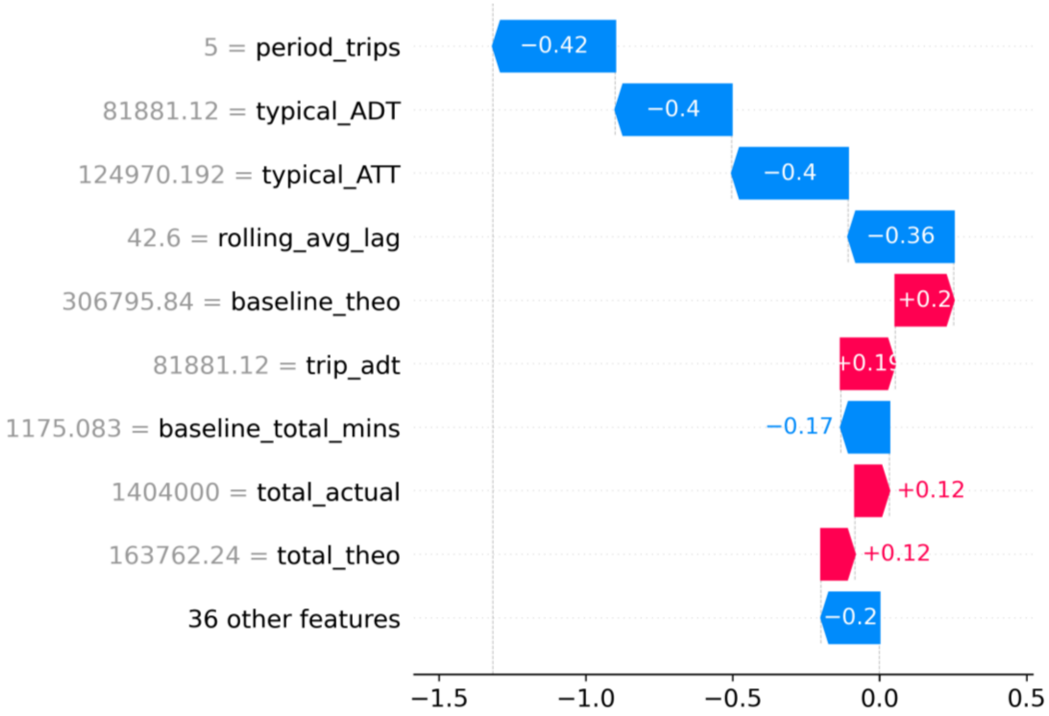
Most important features for churn prediction



Offer should: Instill regularity of visits while addressing recent losses and sudden changes in behavior

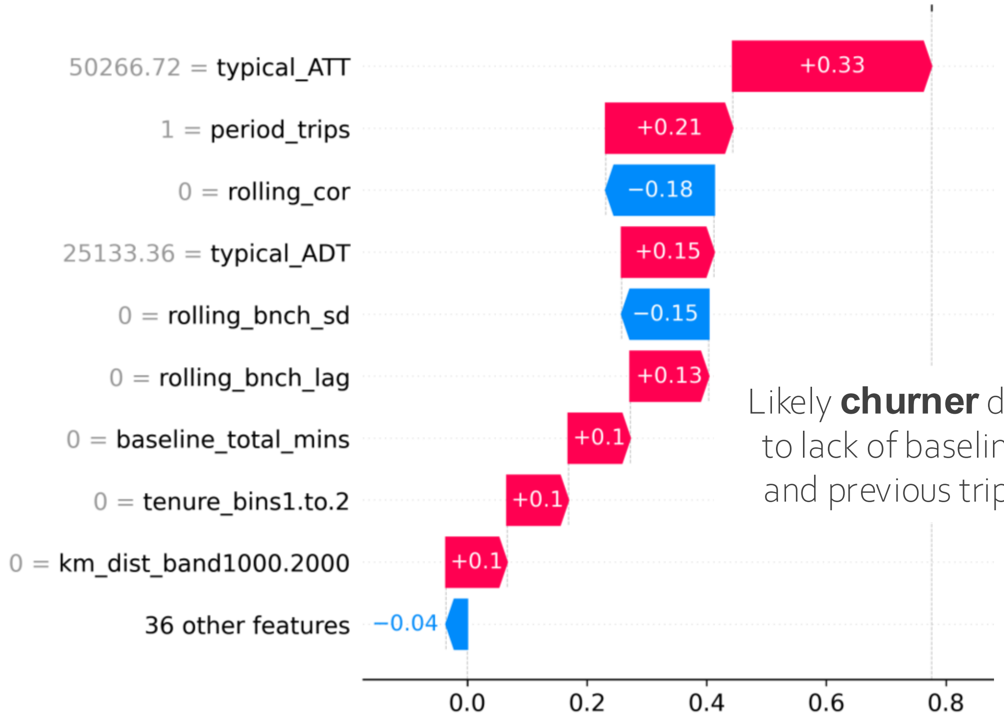
Churners behave very differently in marketable ways

Player 1235: not likely to churn



This player is an engaged, high value player, coming regularly

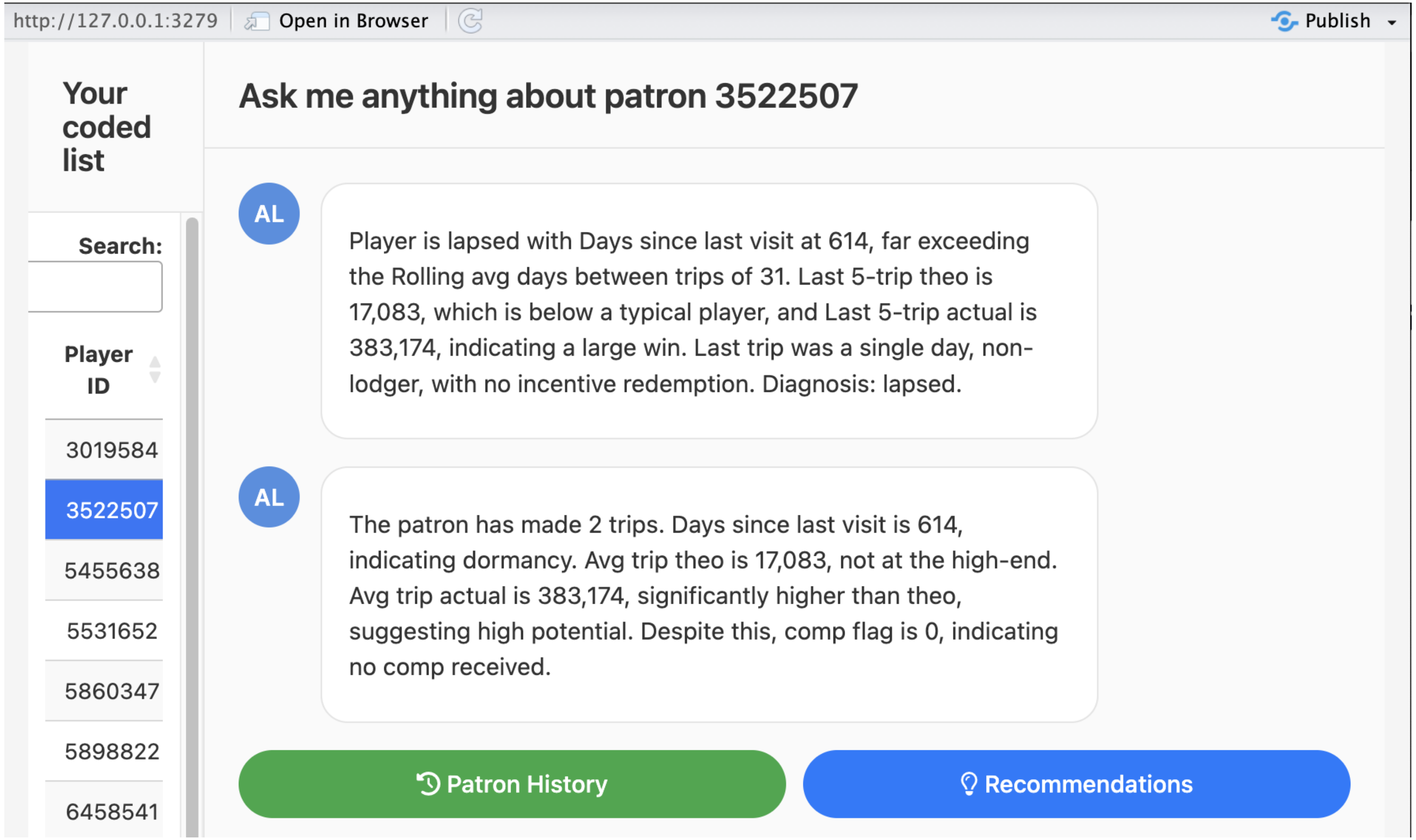
Player 5321: likely to churn



Likely **churner** due to lack of baseline and previous trips

This player is likely new and unknown to us. She has a relatively strong ADT, but not likely to be coming back

We found a way to bridge the gap between analysts and other stakeholders



SECTION THREE

On academia…

So many disciplines are being transformed by AI, and there seems to be a general publication bias towards ML models


Gao et al. Financial Innovation (2024) 10:86
https://doi.org/10.1186/s40854-024-00629-z
Financial Innovation

REVIEW Open Access

Machine learning in business and finance: a literature review and research opportunities

Hanyao Gao[1], Gang
Yucheng Dong[1,4*]

Science
Current Issue First release papers Archive About Submit manuscript

HOME > SCIENCE > VOL. 380, NO. 6650 > AI AND THE TRANSFORMATION OF SOCIAL SCIENCE RESEARCH

PERSPECTIVE | SOCIAL SCIENCE

AI and the transformation of social science research

Careful bias management and

IGOR GROSSMANN, MATTHEW FEINBERG, DAWN C. F

ELSEVIER
Journal of Medicine, Surgery, and Public Health
Volume 3, August 2024, 100099

Enhancing mental health with Artificial Intelligence: Current trends and future prospects

David B. Olawade [a], Ojima Z. Wada [b], Aderonke Odetayo [c], Aanuoluwapo Clement David-Olawade [d], Fiyinfoluwa Asaolu [e], Judith Eberhardt [f]

A new review seems to come out every week titled:

AI research in ____________

Insert whatever subject you like, it is revolutionizing the way we find relationships in our data and draw conclusions...

XAI can level-up your next paper, particularly in disciplines that are not traditionally quantitative

"I don't really know how to program that well and it seems complicated..."

It takes 30 lines of code

```r
#==============================================================================
# XGBOOST MODEL
#==============================================================================
library(tidyverse)
library(xgboost)
library(shapviz)

# read in the data
data_path <- "/Users/dgmana/diff_gaming Dropbox/mana/manas_world/yeehaw_pres/"
prestige <- readr::read_csv(paste0(data_path, "Prestige.csv")) %>%
  dplyr::mutate(
    type = as.factor(type)
  )

# train model
mm <- model.matrix(prestige ~ education + income + type, data = prestige)[, -1]

# holdout split
X_train <- mm[-7, , drop = FALSE]
y_train <- prestige$prestige[-7]

# train xgb model
dtrain <- xgb.DMatrix(X_train, label = y_train)
xgb_model <- xgb.train(list(objective = "reg:squarederror", eta = 0.1, max_depth = 3),
                       data = dtrain, nrounds = 100, verbose = 0)

# shap values
sv <- shapviz(xgb_model, X = X_train, X_pred = X_train)
sv_importance(sv, kind = "bee")
sv_waterfall(sv, row_id = 1)   # only one row in X_pred
```

The idea of XAI could be implemented as a unit or at least explored in various courses to prep students for the AI world

Undergraduate Courses

CSC 3400 Artificial Intelligence

CSC 3400 Artificial Intelligence

3 Hours

Prerequisites: CSC 1120

A survey of algorithms developed for artificial intelligence, with particular emphasis on applications to cognitive science.

BSA 4420 Deep Learning and AI Ethics

BSA 4420 Deep Learning and AI Ethics

3 Hours

Prerequisites: PHY 1620 or DSC 2010 or BSA 2050 and BSA 3620 or permission of instructor.

This course presents an overview of machine learning techniques using deep neural networks. Topics include: supervised, unsupervised and reinforcement learning approaches; classification & regression tasks; generative and discriminative models; recommendation algorithms, natural language processing, computer vision, audio processing. We will also give attention to issues of ethics and society, including bias, transparency and accountability. Cross Listed with DSC 4420 and PHY 4420.

BSA 4250 Predictive Analytics

BSA 4250 Predictive Analytics

3 Hours

Prerequisites: BSA 3620

This course provides students a practical, hands-on learning environment focusing on data mining and predictive analytics to solve business problems. Students will prepare data, create and validate predictive models, and deploy those models to predict future events and uncover hidden patterns of behavior. Students will examine how data analysis technologies can be used to improve decision-making by studying the fundamental principles and techniques of data mining to develop data-analytic thinking.

DSC 2030 Topics in Data Science

DSC 2030 Topics in Data Science

3 Hours

Prerequisites: DSC 1010 and MTH 1151

Topics will be selected from the following: data mining, data/statistical modeling, visualization, data analysis, natural language processing, cloud computing, GIS, machine learning, and other topics as appropriate. Course offerings and topics will appear in the schedule of classes. This course may be repeated for credit for different topics.

BSA 4650 Big Data Analytics

BSA 4650 Big Data Analytics

3 Hours

Prerequisites: BSA 3620

This course introduces students to the fundamental technologies, platforms, and methods that enable Big Data analysis. Students learn how to process Big Data on platforms that can handle the volume, velocity, variety and veracity of Big Data. Students learn how to setup and operate Big Data platforms to complete real world, Big Data analysis tasks that allow them to become comfortable with summarizing and communicating their findings.

BSA 4350 Data Analytics and Visualization

BSA 4350 Data Analytics and Visualization

3 Hours

Prerequisites: BSA 3620 or MKT 3210

This course equips students with the skills to analyze diverse data sets and leverage analytical tools to address business challenges effectively. Students will identify data requirements, assess data quality and completeness using statistical techniques, and prepare and transform data for analysis. Emphasis is placed on mastering data visualization concepts and theory utilizing tools such as Power BI for data preparation, analysis, and visualization. Through hands-on activities, discussions, software demonstrations, and collaborative projects, students will learn to craft compelling visualizations and communicate actionable insights to support data-informed decision-making. Course formerly listed as MIS 4450- Special Topics in Information Systems Management Cross Listed with MKT 4350

ECO 3600 Econometrics

ECO 3600 Econometrics

3 Hours

Prerequisites: MTH 1240 *or* MTH 1150 *and* ECO 2210 *or* ECO 2220.

This course is designed to introduce basic econometric principles as an essential tool for economic analysis. Probability concepts along with simple linear and multiple regression techniques will be the primary focus of the course. Students will become skilled at using computer software as one of the most powerful tools to perform business research and will gain knowledge of research skills that can be applied in their professional careers or graduate studies.

Graduate Courses

BSA 6830 Data and Decision Making

BSA 6830 Data and Decision Making

3 Hours

This course teaches students how to work with different types of data and utilize analytical tools to solve problems. Students will identify data requirements, utilize statistical techniques to evaluate data quality and completeness, prepare data for analysis, and transform data into useful information.

BSA 6820 Predictive Analytics

BSA 6820 Predictive Analytics

3 Hours

This course provides students a practical, hands-on learning environment focusing on data mining and predictive analytics to solve business problems. Students will prepare data, create and validate predictive models, and deploy those models to predict future events and uncover hidden patterns of behavior. Students will examine how data analysis technologies can be used to improve decision-making by studying the fundamental principles and techniques of data mining to develop data-analytic thinking.

BSA 6750 Artificial Intelligence and Machine Learning

BSA 6750 Artificial Intelligence and Machine Learning

3 Hours

Prerequisites: BSA 6650

This course provides an in-depth exploration of artificial intelligence (AI) and machine learning (ML), focusing on techniques and applications relevant to today's data-driven society. Students will learn about the data mining process, data collection, exploration and pre-processing, regression models, decision trees, neural networks, ensemble modeling, document classification, sentiment analysis, correlations, clustering, and the differences between supervised and unsupervised learning. The course also emphasizes hands-on learning and real-world applications.

BSA 6350 Regression Analysis

BSA 6350 Regression Analysis

3 Hours

Prerequisites: BSA 6100

This course offers students an in-depth exploration of regression analysis techniques used for interpreting and analyzing data in various business settings. Students will learn how to prepare data, fit regression models, identify model weaknesses, draw conclusions, and recommend business actions. They will also explore specific techniques such as least squares, model plots, transformations, model selection techniques, linear regression splines, logistic regression, Poisson regression, and Bayesian approaches.

… And of course we can't forget about research methods at the graduate level

Survey Research

AI and XAI can uncover the most impactful components of scales and constructs.

Medical Research

AI detects complex patterns in massive datasets like genomics or imaging, while XAI makes those discoveries transparent and clinically meaningful.

Business Research

XAI helps refine theories of consumer behavior by revealing which factors truly drive decisions

Thank you for having me

differential labs.

Mana Azizsoltani
Head of Data Science

mana@diffgaming.com
+1 (702) 445-3737
https://www.linkedin.com/in/manaazizsoltani/