playtech protect

Risk Identification and Explanation

How to explain the reasons behind a player being risky

To find out more about this research or if you have any suggestions for future topics to be addressed via the Industry Research Brief, please contact the team via Protect@Playtech.com

playtech protect

RG Risk Identification and Explanation

How to explain the reasons behind a player being risky

What is the topic?

In responsible gambling (RG), it is increasingly important to identify player who may engage in problematic gambling if proper interventions are not made at the right time. Over the years, an increased number of tools to identify risky players have been made available to the responsible / safer gambing market place, including Playtech's BetBuddy product suite. Several of these tools draw on machine learning and AI to build logical, mathematical and/or statistical models to identify at-risk players. Are these models self-explanatory? To what extent do they provide an explanation of why a player has been identified as at-risk? Do they require "human in the loop" features to deliver most value?

Why is it important?

Estimating that a given player might be at-risk is only the first step in a targeted intervention framework. If the intrinsic reason(s) behind a player being identified as at-risk are not identified, only generic interactions might possible, whereas tailored, personalised interactions have been shown to be more effective. For example, if it can be identified that a player might be at-risk due to high and sporadic deposit amounts, an interaction recommending the use of deposit limit tools to help manage their depositing behaviour can be made. This is more likely to be effective with this player than an interaction that recommended the use of time limits.

What did the research do?

Extensive researches were carried out to look into various ways to understand why a machine learning based AI model identifies a player as at-risk. There are AI / machine learning models which are known as transparent models or white box models. Explanation of prediction from this type of models is rather simpler compared to explanation of prediction of opaque models, which is often referred as black box model. The question may be raised, if white box models are easy to explain, why do we need to use black box models at all? Because of their simplicity, white box models do not always perform to acceptable standard when put into practice. They often fail to learn from complex multi-dimensional data. One exception is EBM (Explainable Boosting Machine). EBMs, despite being mostly white box model due to the generalized additive approach, can learn well from complex data and typically provide acceptable level of performance (Caruana et al., 2013, 2019).

To identify risky gambling, most of the RG tools employ black box models, built using specialised machine learning algorithms. Each of these black box machine learning models might contain millons of complex mathematical and statistical rules. This research briefing focuses on findings of research done in the machine learning and data science communities around the world, and also findings from Playtech Protect's own research.

Sensitivity: Internal

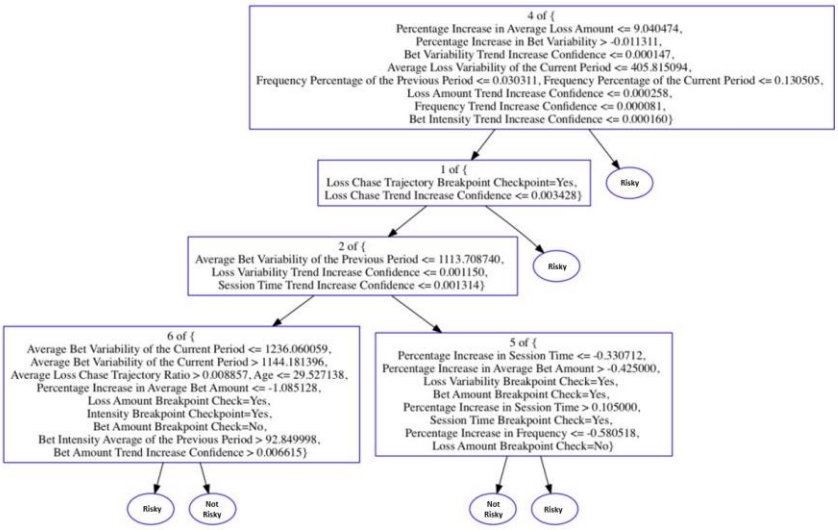
What did the research find?

Researchers around the machine learning communities attempted to discover various ways to interpret the internal working of black box models to some extent. There are popular approaches used in major ML deployments: explanation by simplification, feature relevance and visual explanation.

Explanation by simplification

Explanation by simplification is based on the inputs into the black box model and its outputs. To explain the working of a black box model, researchers (e.g. Hara and Hayashi, 2016; Van Assche and Blockeel, 2007; Zhou and Hooker, 2016), tried to generate one or more simplified models containing sequential or hierarchical decision rules (decision trees). However, for a complex black box model, each single decision tree may take a complex form as we at Playtech Protect found in research with TREPAN tree (Sarkar et al, 2016).

Exhibit 1: simplified TREPAN tree for global explanation of black box model



In exhibit 1, the decision tree is built to understand the working of a black box model using a technique called TREPAN (Craven, 1996). Each node of the decision tree contains rules using multiple behavioural features of players and several thresholds. Obtaining reasons on why a player is identified as at-risk is still difficult despite simplifying the complex black box model into several such decision trees.

Feature relevance

The feature relevance approach attempts to explain a black box model through the contribution of different gambling behavioural features during the model building/learning and prediction process. Feature relevance techniques are applied both at the model level (global level) and at the player level to understand reasoning for individual players being identified as at-risk (local level).

One popular technique is SHapley Additive exPlanation or SHAP (Lundberg and Lee, 2017). Another technique is Local Interpretable Model-agnostic Explanation or LIME (Ribeiro, Singh, Guestrin, 2018),

Sensitivity: Internal

which identifies the degree of influence of behavioural features towards a machine learning model's decision making process. Further work on this was carried out (Datta et al., 2016) which researched measures taken to qualify the degree of influence of a behavioural feature on identifying a gambler as at-risk.

Exhibit 2: Comparison of two technoques – SHAP and LIME for global explanation of black box model

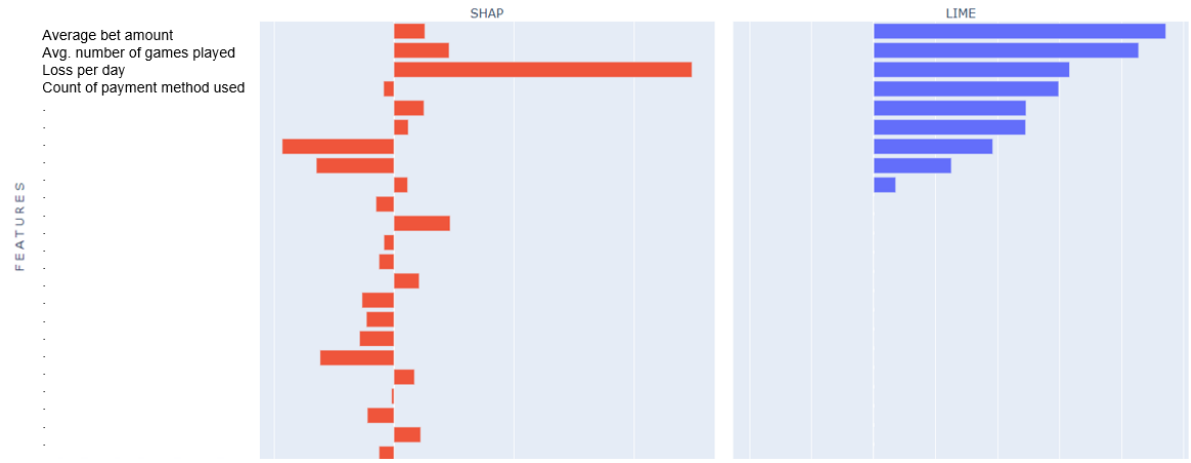


Exhibit 2 depicts the feature relevance explanation as to why a player is identified as at-risk by one of BetBuddy's machine learning models. Two feature relevance explanation methods are applied – SHAP and LIME. While SHAP provides the degree of influence and its direction of influence, LIME only provides the degree of influence. Moreover, there are differences among the reported degree of influences by these two techniques.

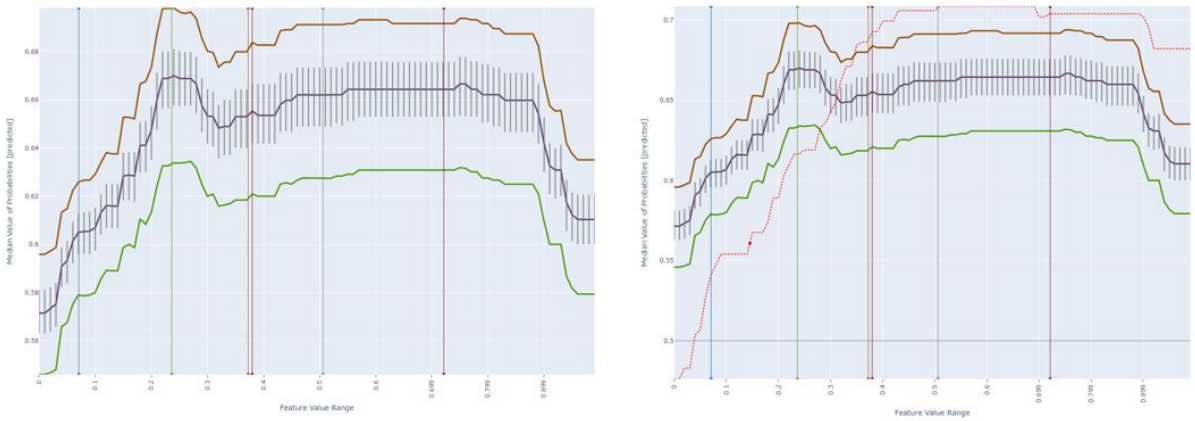
Visual explanation

The visual explanation approach focuses on several 2-dimensional graphs where each graph depicts the behaviour of the model as the values of one or more features are altered. For example, these plots may indicate how risk levels increase or decrease if one of the behavioural feature, say, deposit amount, is increased from its lowest value to its highest value. By examining several visual graphs, it may be possible to understand part of the overall working of a black box model. Several studies have been carried out in this area of explanation (Cortez and Embrechts, 2011,2013; Friedman, 2001; Goldstein et al., 2013).

Playtech carried out research in this area starting with (Percy, Garcez, Dragicevic, Sarkar, 2019) and tried to identify best fit for explanation of decision made by BetBuddy machine learning models.

Sensitivity: Internal

Exhibit 3: Feature risk curve: for global and local level explanation of black box model



Global Level Explanation **Local Level Explanation**

Exhibit 3 contains two graphs, known as feature-risk curves. The graphs show how the probability of being identified as at-risk changes with the values of one of the gambling behavioural features in the model, keeping the other behavioural features unchanged.

In this example, the gambling behavioral feature of percentage of time a player played between midnight and 8 am (night play) is used. The graph on the left depicts model level (global level) explanation, whereas the one on the right depicts player level (local level) explanation. Curves in green colour and maroon colour indicate the 25th percentile and 75th percentile values of the probability, respectively when the value of the behavioural feature is changed from low value to high value along the x-axis. The height of the vertical grey bars denote the confidence intervals between 40th and 60th percentile of probability value.

At the global level, up to approximately 25% of night time play, the model shows a steady increase in the chance of getting identified as at-risk. From that value onwards, till approximately 80%, there is not much significant increase or decrease in the chance of being identified as at-risk due to night time play. From 80% onwards, increased night time play is associated with decreaing chance of being identified as at-risk.

Explanation of why a single player is being identified as at-risk can be obtained from the graph on the right in exhibit 3. Along with the model level explanation, the graph also plotted the red dotted line as the change in probability of being identified as at-risk for that single player due to night play. For this specific player, based on their specific gambling activities, their chance of being identified as at-risk increases until a peak at around 50% of night play, remaining broadly constant from there onwards.

Knowledge of recent changes in that player's night play levels can be combined with this feature risk curve to help understand whether and how much the model might be drawing on night play features to assess them as at-risk.

Sensitivity: Internal

What are the takeaways for safer gambling?

Various researchers have investigated the requirements of model explainability (how the RG machine learning model work?) and interpretability (why does the RG machine learning model predict a player as risky player?) of a black box model. None of the approaches shown fully address the internal logic within black box models. In general, the research in this area aims to build simple, approximate summaries of what most influences model predictions.

From the safer gambling point of view, it is important for the compliance team and customer support team of a gambling operator to know why a player is being identified as at-risk. This will help them have focussed and more effective interactions with those players. Moverover, safer gambling not only requires understanding the reasons for a player to be identified as at-risk, but also requires that these reasons are capable of motivating behavioural change, i.e. being recognised as relevant to their own behaviour by the player and within their direct or indirect control. For example, if three possible reasons are obtained by using one of these explanation techniques, it may happen that only one reason has a credible chance of informing a behavioural change. Safer gambling personnel may need to rely on these reasonings with the caveat that not all possible reasons suggested by the explanation techniques may be actionable in this respect.

Though some conflicts in interpretability may be observed among techniques for explanation, they can be complementary to one another. For a better explanation of causes of at-risk identification, we recommend using multiple techniques, so that results can be compared and contrasted, both to best understand the model and to best identify factors that could inform meaningful interventions.

Due care should also be taken when using AI from an ethical point of view. There may be several ethical areas which may need to be addressed, one of those being bias. If a model is built with a biased dataset, it may lead to one or more behavioral features being provided as a reason for being at-risk for all players, regardless of their playing behaviour.

How can I find out more?

To find out more about this research or if you have any suggestions for future topics to be addressed via the Industry Research Brief, please contact the team via protect@playtech.com.

Sensitivity: Internal

References

Caruana, R., et al. (2013). Accurate Intelligible Models with Pairwise Interactions. https://www.cs.cornell.edu/~yinlou/papers/lou-kdd13.pdf

Caruana, R., et al. (2019). InterpretML: A Unified Framework for Machine Learning Interpretability (https://arxiv.org/pdf/1909.09223.pdf)

Craven, M. W. (1996) Extracting Comprehensible models from trained Neural Networks, PhD Thesis, Computer Science Department, University of Wisconsin, Madison, WI

Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. IEEE Symposium on Security and Privacy (SP). New York, NY: Institute of Electrical and Electronics Engineers, 598–617.

Hara, S., & Hayashi, K. (2016). Making Tree Ensembles Interpretable. Lanzarote, Spain: Proceedings of Machine Learning Research.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, December 2017, NIPS'17. NY, USA: Red HookCurran Associates Inc, 4768–4777.

Percy, C., Garcez, A., Dragicevic, S., & Sarkar, S. (2019). Understanding the Risk Profile of Gambling Behaviour through Machine Learning Predictive Modelling and Explanation. 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada. Available via https://kr2ml.github.io/2019/papers/ and https://kr2ml.github.io/2019/papers/KR2ML_2019_paper_33.pdf

Ribeiro, M. T., Singh, S., & Guestrin, C.. (2018). Anchors: High-precision Model-Agnostic Explanations. New Orleans Riverside, New Orleans: AAAI Press. (https://arxiv.org/pdf/1602.04938v1.pdf)

Sarkar, S., Weyde, T., Garcez, A. D., Slabaugh, G., Dragicevic, S., & Percy, C. (2016). Accuracy and interpretability trade-offs in machine learning applied to safer gambling. Paper presented at NIPS Barcelona (December 2016), published in CEUR Workshop Proceedings, 1773. Available via http://ceur-ws.org/Vol-1773/ or http://ceur-ws.org/Vol-1773/CoCoNIPS_2016_paper10.pdf

Van den Berg, M., & Kuiper, O. (2020). XAI in the Financial Sector. A Conceptual Framework for Explainable AI (XAI), Hogeschool Utrecht, Lectoraat Artificial Intelligence Version 1.1.

Zhou, Y., & Hooker, G. (2016). Interpreting Models via Single Tree Approximation. (https://arxiv.org/pdf/1610.09036.pdf)

Sensitivity: Internal