Accuracy and Interpretability Trade-offs in Machine Learning Applied to Safer Gambling

Sanjoy Sarkar
Department of Computer Science and BetBuddy Ltd.
City, University of London London, UK
Sanjoy Sarkar@city.ac.uk

Tillman Weyde, Artur d'Avila Garcez, Gregory Slabaugh
Department of Computer Science
City, University of London
a.garcez@city.ac.uk

Simo Dragicevic, Chris Percy
BetBuddy, London, UK
simo@bet-buddy.com, cwspercy@gmail.com

Abstract

Responsible gambling is an area of research and industry which seeks to understand the pathways to harm from gambling and implement programmes to reduce or prevent harm that gambling might cause. There is a growing body of research that has used gambling behavioural data to model and predict harmful gambling, and the industry is showing increasing interest in technologies that can help gambling operators to better predict harm and prevent it through appropriate interventions. However, industry surveys and feedback clearly indicate that in order to enable wider adoption of such data-driven methods, industry and policy makers require a greater understanding of how machine learning methods make these predictions.

In this paper, we make use of the TREPAN algorithm for extracting decision trees from Neural Networks and Random Forests. We present the first comparative evaluation of predictive performance and tree properties for extracted trees, which is also the first comparative evaluation of knowledge extraction for safer gambling. Results indicate that TREPAN extracts better performing trees than direct learning of decision trees from the data. Overall, trees extracted with TREPAN from different models offer a good compromise between prediction accuracy and interpretability. TREPAN can produce decision trees with extended tests rules of different forms, so that interpretability depends on multiple factors. We present detailed results and a discussion of the trade-offs with regard to performance and interpretability and use in the gambling industry.

1 Introduction

The application of machine learning to understand gambling pathways to harm and addiction is a new and growing field of study. Account-based gambling, whether via Internet or retail channels, whilst traditionally used for marketing purposes, has revolutionized this field of study due to the amount of data available to identify early warning signs of potentially harmful behaviour [7]. Such data was previously anonymous or unregistered, and not attributable to an individual player. However, the

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

quantity of data simultaneously opens up questions of how best to interpret the data and its results: specifically, how to transform raw gambling session data into meaningful, descriptive variables, called behavioural markers, and how to relate those descriptive variables to an individual who is potentially at risk of harm or addiction.

There are two important benefits of being able to predict potential harm in gambling behaviours. The first is improved player protection. By identifying individuals whose play pattern approximates those who have previously experienced harm, the gambling operator can choose to share information or advice with the player that may support healthy engagement with the gambling platform. Alternatively, the operator may choose to restrict marketing activity or platform activities for that player for a certain period of time. For this to happen effectively, interpretability of results is important. The second benefit are more stable, long-term revenue flows to gambling operators, since gamblers that might use their platform less intensively than before may do so with greater security and satisfaction.

Whilst the current machine learning methods offer good prediction performance, their effectiveness will be limited by the machine's inability to explain its decisions and actions to users. Explainable machine learning will be essential if users are to understand, appropriately trust, and effectively manage this incoming generation of artificially intelligent partners [8]. In the context of gambling, whilst machine learning algorithms have demonstrated early promise by predicting potentially harmful gamblers [10, 12], the industry uptake of such systems will primarily be dependent upon the regulators' and gambling operators' ability to understand and effectively use them. In an effort to overcome these challenges, previous research [15] has applied the TREPAN knowledge extraction method to neural networks in an effort to understand aspects of harmful gambling behaviour e.g. which kinds of profiles fit into problematic gambling, which attributes explain players who have such profiles? Such questions are motivated by industry insight from a responsible gambling conference in Vancouver (New Horizons in Responsible Gambling, 2016) [9] and an industry seminar in London (Responsible Gambling Algorithms Roundtable, 2016) [8], in which gambling operators, treatment providers and public policy officials set out the need for effective interpretation of such complex machine learning algorithms.

In this paper, we apply TREPAN to random forests and neural networks and offer the first comparative analysis of extracted trees from different models with different parameters regarding their accuracies and interpretability. This is also the first comparative study of knowledge extraction for safer gambling. Results indicate that TREPAN is a useful technique to aid interpretation of random forest and neural networks, leading to improved performance compared to standard decision trees. Different models, extraction parameters and tree types lead to varied loss of accuracy and degrees of interpretability.

The remainder of this paper is organised as follows: Section 2 discusses the related work in the application of machine learning to understand and interpret gambling behaviour. Section 3 describes the extraction and comparison methodology, including the changes to TREPAN to enable application to both random forests and neural networks. Section 4 presents the results comparing model accuracy and fidelity to the original random forest and neural network and between different types of extracted trees. Section 5 discusses the interpretability of our empirical results, and concludes the need for further research of understanding and measuring algorithm interpretation.

2 Related Work

2.1 Predicting Harm in Gambling

Machine learning algorithms have only recently been applied to this field of study as a way of predicting potentially harmful gambling [10, 12]. In [10], data obtained from the gambling operator International Game Technology PLC (IGT) was used to describe Internet gambling self-excluders in terms of their demographic and behavioural characteristics. Data analysis approaches and methods for improving the accuracy of predicting self-excluders are developed by hand towards inferred behaviour models. Supervised machine learning models were evaluated in [14] in the context of predicting which gamblers could be at risk of problem gambling. Their results suggest useful but general methods and techniques for building models that can predict gamblers at risk of harm.

Building on the work from the live action sports betting dataset available from the Division on Addiction public domain, in [12] nine supervised learning methods were assessed at identifying disordered Internet sports gamblers. The supervised learning methods include logistic regression and

other regularized general linear models (GLM), neural networks, support vector machines (SVM) and random forests. The results ranged from 62% to 67% with a random forest model reaching the highest prediction accuracy on a test set. A key finding from [10] was that the random forest technique performed the best in overall model accuracy (87%). The test set accuracy of the logistic regression model was the lowest (72%), with Bayesian networks in second and neural networks in between.

However, a major limitation of this study was the lack of interpretability of the models. The random forests were very difficult to interpret, consisting of 200 binary decision trees of unlimited depth. The neural network used was perceptron with a a single hidden layer, with 33 inputs, 17 hidden neurons and 2 outputs, one for self-excluding players and the other for the control group players. With more than 500 weights, the neural network was also a black box. The Bayesian network, which used the K2 algorithm, included hundreds of separately defined conditional probabilities and also failed to instigate useful insight about the problem when shown to industry experts.

2.2 Industry Need for Knowledge Extraction and Explainable Prediction Models

As reported in [15], we polled the audience at a related presentation at the 2016 New Horizons in Responsible Gambling conference to explore the importance of knowledge extraction and algorithm interpretability. Respondents were asked whether they would prefer a responsible gambling assessment algorithm that provided a 90% accurate assessment of problem gambling risk that they could not unpack or understand, or a model that provided a 75% accurate assessment that was fully interpretable and accountable. Only 20% chose the more accurate model, with 70% preferring to sacrifice 15 percentage points of accuracy for greater interpretability (10% were uncertain or felt it depended on the circumstances).

In a further exercise undertaken at the Responsible Gambling Algorithms roundtable event held in London in 2016 [8], we asked senior industry stakeholders for their views on the importance of understanding and interpreting algorithms. Senior executives and experts, including participants drawn from gambling operators as well as representatives from treatment providers and the UK Gambling Commission and the UK Responsible Gambling Strategy Board, were asked if accuracy of algorithms for recognising gamblers at risk was considered a second priority compared to the need for understanding them. The unanimous consensus from the participants expressed a preference for a more understandable algorithm over a more accurate one. For example, Dirk Hansen, CEO of GamCare, the UK's largest problem gambling treatment provider, stressed the value of interpretability, especially as this can be an advantage when providing treatment as the counsellor has specific and relevant behavioural indicators to discuss. From a regulatory perspective, Paul Hope, Director from the UK Gambling Commission, the industry regulator, stated that greater model understanding would be a higher priority compared with greater accuracy.

2.3 Knowledge Extraction from Neural Networks

This paper focuses on knowledge extraction by using random forests and artificial neural networks and TREPAN on a new IGT dataset to not only predict, but also describe, self-excluders through knowledge extraction. Previously, in [15] a variant of TREPAN was applied to the neural network model trained on gambling data in [10] to produce compact, human-readable logic rules efficiently. To the best of our knowledge, this was the first industrial-strength application of knowledge extraction from neural networks, which otherwise are black boxes and unable to provide the explanatory insights which are required in this area of application. The research demonstrated that through knowledge extraction one can explore and validate the kinds of behavioural and demographic profiles that best predict self-exclusion, while developing a machine learning approach with greater potential for adoption by industry and treatment providers. Experimental results in [15] reported that the rules extracted achieved high fidelity (87%) to the trained neural network while maintaining competitive accuracy (1 percentage point reduction in overall accuracy) and providing useful insight to domain experts in responsible gambling. This raises the real possibility of implementing algorithms for responsible gambling that offer both high accuracy and high transparency and interpretability. However, a limitation of this research is that it did not apply TREPAN to other machine learning methods, notably random forests, which were the most accurate model in [10].

3 Method

Gambling Data and Data Preparation The IGT dataset is based on gambling behavioural data made available by IGT collected from 4th December 2014 to 30th June 2016 from the regulated Internet gambling jurisdiction of Ontario, Canada. The sample data for the prediction model development and testing was based on Internet casino play and comprised 13,615 control group players and 449 self-excluders who self-excluded for at least six months. Self-exclusion is only a secondary indicator of disordered gambling behaviour, but specifically voluntarily exclusion from gambling platforms for significant periods of time has been previously used a dependent variable for developing models to predict potential harm in gambling [10].

The attributes of these players' raw activity data are a de-identified player unique ID, date of play, start time and end time of play sessions, type of game, game name, bet amount, and win amount. A number of behavioural markers are extracted that represent known aspects of risk, such as how much time gamblers spend on-line or how much they bet and how this evolves over time. For details of how the behavioural markers were generated, please see [10]. In addition to the 33 features used in [10], an additional 17 features were engineered and added to this data set to model additional behavioural markers around loss behaviours, such as increasing losses, increasing variation in the size of losses, and increasing loss chasing behaviours.

After processing the dataset contained 14,112 samples, each comprising 50 features per player. For each set, standard descriptive statistics measures are use to identify nature of data distribution and variance. A small number of samples with missing information was removed. This led to above mentioned of 13,615 control group player's samples and 449 self-excluder samples.

For the purpose of model building, a balanced training dataset was created by generating artificial samples for the minority class (self-excluder) using SMOTE [3].

Our approach for understanding gambling behaviour through machine learning is composed of three steps: gambling data preparation, models building using random forest and neural network algorithm, and knowledge extraction using TREPAN.

Model Development For the purpose of model development we a combination of random subsampling the control group and oversampling the self-excluders using SMOTE [3] to create a dataset with 1685 data points in each class.

The Random Forest models were built with forest size of 200 binary double trees, with unlimited depth. The neural network models contained 38 hidden nodes and 2 output nodes. A learning rate of 0.2 and momentum of 0.2 were used. The values of these hyper parameters were determined in a grid search. For comparison we also trained a standard decision tree with unlimited depth and number of leaves, splitting by the Gini criterion using the *scikit-learn* package in Python. The models are evaluated in 10-fold cross validation.

Knowledge extraction using TREPAN The original motivation of Craven's work in [4] was to represent a neural network model in a tree structure which could be more interpretable than a neural network classification model. This was in the context of a wider interest in knowledge extraction from neural networks [1,2], of which TREPAN has the advantage of being applicable to any oracle. In this work, the motivation was to apply Craven's method to a random forest model in addition to a neural network model and to explore the use of different types of decision rules in the generated trees. TREPAN generates decision rules of type *M of N* , *N of N*, *1 of N*, or *1 of 1*. In an *M of N* configuration, a tree node contains N distinct tests. If out of these N tests, M tests are satisfied, the tree will take one decision path, otherwise, the other path. *N of N* is a special case where $N = M$, creating a logical conjunction of tests. *1 of N* is another special case with $M = 1$, creating a logical disjunction of tests. *1 of 1* creates a standard decision tree with a single test per node. The type of rule is can be combined with different tree sizes in the tree generation process.

4 Experimental Results

Table 1 shows results of the random forest, neural network and decision tree models. In terms of classification accuracy, the results confirmed earlier research [10,15] that showed that random forests performed better than neural networks on this type of data and that both methods outperform a

Table 1: The structure and performance or random forest and neural network models.

Model Structure	Random Forest	Neural Network	Decision Tree
No of Trees	200		1
Tree Height	16 - 28		
Tree Leaves	343 - 432		
No of nodes in Hidden Layer		38	
Area under the ROC Curve	0.95	0.91	0.62
Accuracy	90%	84%	76%
True Positive Rate	85%	80%	80%
True Negative Rate	93%	88%	73%

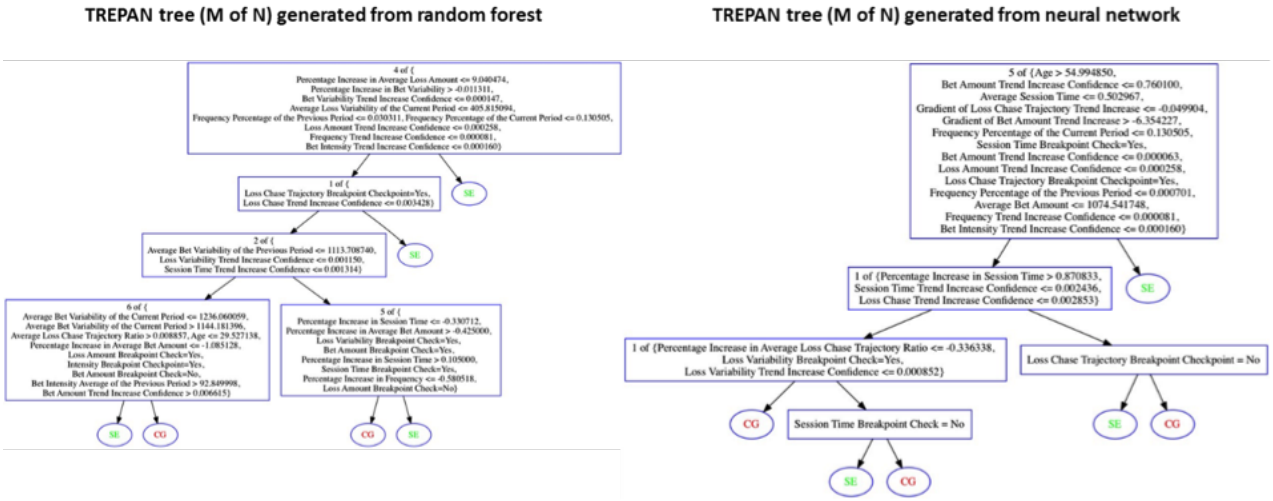

Figure 1: Decision trees generated with TREPAN for both a random forest and a neural network. The maximum number of internal nodes was set to 10.

standard decision tree with cross-validated accuracies of 90%, 84% and 76% respectively. We also tested neural networks with 2 and 3 hidden layers (with the same learning method as described above), but they did not improve results on this dataset.

Figure 1 depicts two trees generated by TREPAN.[1] CG at a leaf represents the control group class and SE represents the self-excluder class. For both trees, the maximum number of internal nodes is 10. It can be observed that considerable complexity is shown in the $MofN$ configuration the several internal nodes in each tree, especially for the larger trees.

Table 2 shows the performance of TREPAN generated trees for three different values of maximum number of internal nodes when applied to the two models. The fidelity of the model denotes the agreement between the TREPAN model and the original model. The accuracy of TREPAN model denotes the agreement with the training dataset used for model development. As expected, the overall accuracy of the TREPAN trees is reduced compared to the original models' performance for both models. The random forest was more accurate than the neural network, ranging from a difference of 4% to 1%, depending on the internal node size. The loss of accuracy is however lower for the trees extracted from the neural network (2–4%) then from the random forest (6–7%). The TREPAN trees extracted from the neural network showed higher fidelity to the original model (85%–87%), than those extracted from random forests (80%–81%). The internal node size of the TREPAN had little impact on accuracy or fidelity to the original models for random forests or neural networks.

The statistics of the tree and internal node structure that follow in Table 2 (M of N, feature counts, internal nodes, leaves, depth, feature count) show some differences of interpretability. The TREPAN trees where $max.internalnodes = 20$, when compared to those with $max.internalnodes = 5$, typically had more *M of N* nodes (5 and 4 v. 4 and 1 for random forests and neural networks), more

[1] Images generated with `http://www.graphviz.org/`.

Table 2: Descriptive summary of TREPAN decision trees generated for both random forest and neural networks.

Model Size	**Max. Int. Nodes = 5**		**Max. Int. Nodes = 10**		**Max. Int. Nodes = 20**	
Model Type	**RF**	**NN**	**RF**	**NN**	**RF**	**NN**
Accuracy	84%	80%	83%	82%	83%	80%
Fidelity	84%	87%	83%	87%	83%	85%
M of N Nodes	4	1	5	3	5	4
feature count in M of N	8,2,3,8	11	8,2,3,9,8	14,3,3	9,2,3,10,8	11,4,17,10
Internal Nodes	4	1	5	5	10	10
Leaves	5	2	6	6	11	11
Depth	4	1	4	4	7	8
Feature Count	21	11	30	22	37	48

Table 3: Comparison of results for both random forest and neural networks using four different TREPAN tree structures

Type	**Size of tree**	**Random Forest**		**Neural Network**	
		Accuracy	**Fidelity**	**Accuracy**	**Fidelity**
M of N	*Max internal node size = 5*	84%	84%	80%	87%
	Max internal node size = 10	83%	83%	82%	87%
	Max internal node size = 20	83%	83%	80%	85%
N of N	*Max internal node size = 5*	52%	52%	82%	78%
	Max internal node size = 10	69%	69%	79%	79%
	Max internal node size = 20	70%	70%	83%	81%
1 of N	*Max internal node size = 5*	51%	51%	83%	79%
	Max internal node size = 10	75%	75%	84%	79%
	Max internal node size = 20	78%	78%	85%	80%
1 of 1	*Max internal node size = 5*	51%	51%	47%	54%
	Max internal node size = 10	57%	57%	75%	70%
	Max internal node size = 20	68%	68%	75%	73%

leaves (11 and 11 v. 5 and 2), and more features used (37 and 48 v. 21 and 11). However, this more complex structure did not translate into a notable improvement performance.

In addition to the default $MofN$ type trees mentioned above with arbitrary M, we also generated TREPAN trees using $NofN$, $1ofN$, and $1of1$ structures to compare model performance, as described in table 3. The results show that for the random forest, the $NofN$ and $1ofN$ lost much accuracy compared to $MofN$, while there was no such loss for the neural network. This is an interesting result, as the $MofN$ tress were reported as hard to interpret in initial expert feedback, particularly with with large values of N and M. From a logical perspective, the $MofN$ nodes represents a much more complex test rule than $1ofN$ or $NofN$, which correspond to a logical rule with only N components in a disjunction or conjunction. The $1ofN$ case produced the best overall results for all tree types extracted from the neural network, including the single most accurate extracted tree (85%, for max. internal node = 20), which is event more accurate than the neural network model itself.

5 Discussion

The accuracy of random forest is throughout higher than that of the neural network, but irrespective of their performances, both the random forest and neural networks models are still black boxes to the end user. The decision tree generated directly form the data performs much worse than the more complex model, which confirmed the motivation of this study.

It was observed that increasing or reducing maximum number of internal nodes had little effect on the TREPAN model performance against the original model (random forest or neural network). This is an interesting result because the main motivation behind generating TREPAN model trees is to enhance the interpretation of a complex model via a simple logical rule representation. A tree with fewer nodes delivers generally a simpler tree structure with better interpretability. For example, our 200 tree

random forest model would represent 3,000 pages of A4, which makes is practically uninterpretable to humans. The smaller generated trees fit on a single page of A4, whilst reducing accuracy by a relatively small amount. The best extracted tree had 5 percentage points lower accuracy than the best model overall (random forest), which is well below the 15 point loss that was considered acceptable by 80% of respondents in return for full model transparency in [15].

However, as complex test rules are created with in the nodes, it becomes harder interpret models. With $MofN$ nodes, the accuracy of a TREPAN tree generated from a random forest model is always higher than that of the neural network model. This indicates that the TREPAN generated trees represent the structure of a neural network more accurately compared to a random forest model. A possible explanation for this that the operation of neural network nodes (summation and thresholding with a non-linear function) can be more easily represented with $MofN$ nodes. If M is between 1 and M there is a combinatorial explosion of possible test results that leads to different decisions, which makes the interpretation of a decision node a complex undertaking. However, for $NofN$ and $1ofN$ the neural network generated more accurate trees. This is an interesting result, as the $MofN$ tress were reported as hard to interpret in initial expert feedback, particularly with with large values of N and M. From a logical perspective, the $MofN$ nodes represents a much more complex test rule than $1ofN$ or $NofN$, which correspond to a logical rule with only N components in a disjunction or conjunction. This makes particularly the use of $1ofN$ trees generated from a neural network attractive, which are highly accurate, robust against tree size, and easier to interpret.

Future research should be aimed at extending the TREPAN algorithm to find new methods to optimise the tree structures to aid interpretability. This entails further research assessing the value and boundary conditions of the two parameters M and N for human interpretability.

From a user perspective, the internal structure of nodes and the shape of a tree are factors in addition to the size of the tree and number of features that influence ability to interpret or comprehend a decision tree model generated. We hypothesise that that limits of working memory will introduce a non-linearity in the ease of interpretation depending on the number of nodes and the number and type of tests within a node. Also, the congruence of decision tests with familiar concepts is likely to influence interpretability [11] and the usefulness for practitioners such as therapists or in automatically generated personalised messages to the user. In addition to optimising the tree structure, improved visualizations may aid human interpretability further. A deeper understanding of the the interpretability will require further research. Based on the current results, a combination of constraints and selection by expert practitioners can be a practical solutions. Ultimately the interpretability vs. accuracy trade-off will be for key stakeholders, such as regulators, scientific community, and industry to decide upon.

6 Conclusions and Future Work

Responsible gambling can benefit from machine learning models to recognise potentially harmful gambling behaviour. The industry does however demand models that are interpretable for professionals and can provide information to affected users. To fulfil this demand we have conducted the first comparative study of different types of decision trees extracted from neural network and a random forest models with TREPAN for safer gambling.

For a complex machine learning model, TREPAN generates relatively decision trees that provide an approximation of the complex models and lend themselves to human interpretation. We evaluated the extraction of small decision trees from neural network and random forest models with different parametrisations with regard to different metrics of classification performance and with regard to the properties of the resulting trees.

The results show that the loss of accuracy can be kept relatively small (between 0 and 7 percentage points), even for small trees. The complexity of the trees generated by TREPAN depends on the number of nodes and the structure of the rules in the decision nodes, where considerable complexity can occur. Although random forests make the most accurate predictions, the best performing decision tree was generated from a neural network. This was also using a simpler form of rules than other high performing trees and seems therefore to offer the best trade-off in this study between accuracy and interpretability.

However, complex rules inside decision nodes mean that the tree size is not the only metric relevant to interpretability, and further empirical work is needed to understand better what determines interpretability for a user. From an industrial assessment perspective, we propose a further study to assess the interpretability of TREPAN trees by domain experts and potentially for end users which can in turn lead to improved knowledge extraction methods.

References

[1] R. Andrews, J. Diederich, and A. Tickle. Survey and critique of techniques for extracting rules from trained artificial neural networks, Knowledge Based Systems, 8(6), 373-389, 1995

[2] Garcez, A. S. D., Broda, K., and Gabbay, D. (2001). Symbolic knowledge extraction from trained neural networks: A sound approach. Artificial Intelligence, 125(1-2), 155–207.

[3] Bowyer, K., Chawla, N., and Hall, L., Kegelmeyer P. (2002). SMOTE: Synthetic Minority Over sampling Technique. Journal Of Artificial Intelligence Research, 16, pages 321-357.

[4] Craven, M., and Shavlik, J. (1996). Extracting Tree-Structured Representations of Trained Networks. Advances in Neural Information Processing Systems, 37–45.

[5] DARPA call *Explainable AI*: http://www.darpa.mil/program/explainable-artificial-intelligence

[6] Franca, M. V. M., Zaverucha, G., and Garcez, A. S. D. (2014). Fast relational learning using bottom clause propositionalization with artificial neural networks. Machine Learning, 94(1), 81–104.

[7] Gainsbury, S., Wood, R. (2011). Internet gambling policy in critical comparative perspective: The effectiveness of existing regulatory frameworks. International Gambling Studies,11(3), 309-32.

[8] ROUNDTABLE: http://www.bet-buddy.com/media/1190/responsible-gambling-algorithms-roundtable-1-august-2016-final.pdf

[9] Percy, C. (2016). Can 'BlackBox' responsible gambling algorithms be understood by users? A real-world example. New Horizons in Responsible Gambling conference paper, Vancouver, February 2016.

[10] Percy, C., Franca, N., Dragičević, S., and Garcez, A. S. D. (2016). Predicting online gambling self-exclusion: an analysis of the performance of supervised machine learning models. International Gambling Studies. DOI:10.1080/14459795.2016.1151913.

[11] Besold, T., Muggleton, S., Schmid, U., Tamaddoni-Nezhad, A., and Zeller, C. How does Predicate Invention affect Human Comprehensibility?. In Proceedings of the 26th International Conference on Inductive Logic Programming (ILP 2016, Sept. 4th-6th, London). Springer, Accepted.

[12] Philander, K. S. (2013). Identifying high risk online gamblers: a comparison of data mining procedures. International Gambling Studies. DOI: 10.1080/14459795.2013.841721

[13] Karim, A. and Zhou, S. X-TREPAN: a multi class regression and adapted extraction of comprehensible decision tree in artificial neural networks. arXiv:1508.07551, Aug, 2015.

[14] T. Schellinck and T. Schrans (2011). Intelligent design: How to model gambler risk assessment by using loyalty tracking data. Journal of Gambling Issues: 26(1), 51-68.

[15] Chris Percy, Artur S. d'Avila Garcez, Simo Dragicevic, Manoel V. M. França, Greg G. Slabaugh, Tillman Weyde: The Need for Knowledge Extraction: Understanding Harmful Gambling Behavior with Neural Networks. ECAI 2016: 974-981