Predicting Cross-Gaming Propensity Using E-CHAID Analysis

Eunju Suh, Ph.D.
Matt Alhaery

Introduction

Data Mining in the Gaming Industry

As more countries and regions are considering legalizing and/or expanding gaming in their respective markets, many casinos in the global casino market are facing increased competition in player acquisition and retention, as well as revenue generation. Hence, customer relationship management (CRM) has become critical for a profitable relationship with customers. One way in which marketers can effectively compete is mining customer data. Mining and analyzing customer data helps marketers better understand customer behavior and predict specific behaviors. This, in turn, will enable them to identify prospects, segment customers, target specific segments, and inevitably optimize the available marketing resources. For these reasons, big data and predictive analytics have become increasingly important across many industries. While some casinos have embraced data analytics and mined customer data for data driven-marketing (Experfy Insights, 2014), to the best knowledge of the authors of this article, there has been relatively little effort in the gaming literature to discuss the application of data mining methods to the prediction of customer behavior.

Data mining is generally defined as the process of discovering meaningful patterns, relationships and associations hidden in large data sets by examining and modeling the data (Chung & Gray, 1999; SPSS Inc., an IBM Company, 2010; Peacock, 1998). The application of data mining techniques to customer data has gained increasing attention and popularity from marketers in various industries, especially in the telecommunication and banking sectors. The use of data mining methods in the marketing field has helped marketers target specific groups of customers or individuals and customize their marketing offers (McCarty & Hastak, 2007). Data mining is also used to classify customers who are likely to respond to specific marketing promotions such as direct-mail offers and to identify cross-sell and up-sell opportunities.

In the gaming industry, a few anecdotal case studies claim the application of data mining techniques for the analysis of customer data. For example, it was reported that Las Vegas Sands selected SAS, a business intelligence software vendor, for the analysis of the vast amount of customer-related data collected from its multiple gaming facilities (Woodie, 2011). Another company, Harrah's Entertainment Inc., [Caesars Entertainment Corporation] was also reported for its use of data mining techniques and predictive analytics (SAS, 2010). In the academic field, the Division on Addictions at the Cambridge Health Alliance, a Harvard Medical School Teaching Affiliate, has partnered

Eunju Suh. Ph.D.*
Associate Professor
The Chaplin School
of Hospitality & Tourism
Management
Florida International
University
suhe@fiu.edu
*Corresponding Author

Matt Alhaery
Independent Consultant
alhaery@yahoo.com

with *bwin*, one of the major online gaming companies, for collaborative research on the problem gambling detection and responsible gambling initiatives (Burton, 2008). These collaborative research efforts are based on the tracked gaming data of individual online bettors, preventing self-reported biases and recall issues which are typically associated with the self-report questionnaire approach (Burton, 2008). Despite these claims, there has been very little discussion regarding the application of data mining techniques in the gaming industry. Additionally, the authors' experience in the gaming industry confirms the limited application of statistical and data mining methods in customer classification and behavior prediction.

Cross-Gaming Prediction

One of the areas where data mining is potentially useful is identifying casino patrons who are likely to play both slots and table games, also known as cross-gaming (Suh & Alhaery, 2014). While it is a common practice in the gaming industry to divide customers into either slot or table game players and market to them accordingly, some patrons may play both slots and table games or have the potential to play both (Suh & Alhaery, 2014). Casino operators often try to promote table games to slot players and vice versa in the hopes of increasing overall gaming revenues (Brokopp, 2008; Green, 2010; Fleming, 2006; Pollard, 2008). This is similar to cross-selling practices in many industries. For example, casino marketers offer match play coupons for table games and free play coupons for slots. These offers are typically designed not only to increase gaming volumes, but also to encourage players to try new or different games (Lucas, 2004; Suh & Alhaery, 2014). Other examples of cross-gaming promotion include an online gaming company cross-selling blackjack to its online poker players (Fleming, 2006; Pollard, 2008); a game manufacturer introducing a hybrid game which features both video-poker and video-slot components (Brokopp, 2008; IGT, 2011); and a company announcing the development of a device to facilitate the exchange of cash-out vouchers from slot machines for table game chips (e.g., Casino Enterprise Management, 2010; FutureLogic, 2012). Additionally, a recent study that examined the prediction of cross-gaming propensity (Suh & Alhaery, 2014) showed the possibility of generating incremental gaming revenues from cross-gaming prospects. In their study, cross-gamers in most areas exhibited greater spending on gaming and higher visit frequency in comparison to non-cross-gamers.

While casino marketers strive to generate additional gaming revenues through cross-gaming promotions, the typical approach in most casinos is simply sending coupons and play incentives to all casino patrons regardless of patrons' cross-gaming propensities (Suh & Alhaery, 2014). The problem with this approach is that marketing dollars are spent on contacting patrons who are unlikely to respond to the cross-gaming promotions and have no intention to try other types of games. To minimize these inefficient efforts and marketing funds on contacting all casino patrons, a predictive model to identify potential cross-gamers is critical. Despite the potential to generate additional gaming revenues and save marketing costs through the application of predictive analytics, there has been very little discussion on predicting cross-gaming behavior. Hence, this study applies a data mining method in the prediction of cross-gaming propensity using the real-world gaming data. This study would provide a better understanding of casino patrons' cross-gaming behavior and the application of a data mining method for predicting patrons' cross-gaming propensity. Additionally, the derived prediction model provides the estimate of cross-gaming propensity for every patron in the sample population. This estimate can be used to identify which patrons are likely to become cross-gamers. Using the estimates, casino marketers can develop more efficient and cost-effective marketing actions to target only the likely cross-gamers and improve cross-gamer conversion rates. This, in turn, will enable casino marketers to reduce the marking expenses in contacting the wrong customers and improve gaming revenues.

Literature Review

Traditional Approach to Player Value Assessment

In the field of Customer Relationship Management (CRM), RFM analysis is an important marketing tool. RFM stands for recency, frequency, and monetary value. Recency indicates the time interval since a customer's most recent purchase; frequency represents the number of times that a customer made a purchase within a specific period of time; and monetary value represents the amount of money spent on purchasing during a certain time period (Olson, Cao, Gu, & Lee, 2009; IBM Knowledge Center, 2011b). These RFM variables are commonly used in direct marketing for distinguishing between likely respondents and non-respondents (Olson et al., 2009; IBM Knowledge Center, 2011b).

It is a common practice in the gaming industry to classify players into different segments (tiers) based on the RFM-related variables for marketing purposes (Lucas & Kilby, 2008; Suh & Alhaery, 2014). The theoretical-win (t-win) value of a player, for example, can be considered as the monetary component of RFM. T-win for table gaming is calculated by multiplying the following numbers: average bet, hours played, decisions per hour and house advantage, while t-win for slot gaming is the product of the amount wagered and house advantage (Lucas & Kilby, 2008). Using the t-win values, the Average Daily Theoretical (ADT) and the Average Trip Theoretical (ATT) can be calculated. ADT represents the average dollar amount that a player is expected to lose per day for a certain gaming activity, and ATT represents the same per trip. Typically, ADT is used for a casino serving mainly local residents, whereas ATT is used for a destination casino serving mainly tourists (Lucas & Kilby, 2008). As the ATT and ADT calculations account for the casino patron's visit frequency during a given time period, it can be said that ATT and ADT incorporate the frequency component of RFM. With respect to R (Recency), casino marketers often review the time since the casino patron's play day to decide to whom direct-mail offers should be sent (Lucas & Kilby, 2008). According to the experiences of the authors of this article in the gaming industry, the RFM-related variables are also used to treat one-time or first-time visitors differently from established patrons as well as to market to active players differently than inactive ones.

Criticism of the RFM and ATT/ADT Approach

The popularity of RMF and its wide acceptance in various fields are often attributed to the method's simplicity and easiness to understand and use (Mutyala, 2011; IBM Knowledge Center, 2011b). However, researchers have questioned the performance of RFM analysis in classifying subjects into distinct groups and suggested a data mining technique such as Chi-squared Automatic Interaction Detection (CHAID) as an alternative to RFM (e.g., McCarty & Hastak, 2007; Olson et al., 2009). One of the criticisms regarding the simple RFM approach is that it is not applicable to new customers or prospects given that there is no transactional data about them (McCarty & Hastak, 2007). The same criticism can be applied to the ADT and ATT measures for player evaluation that are commonly used in the gaming industry. While ADT and ATT are simple to use, they are observed values. Hence, they do not represent the potential gaming worth of new customers with little to no play data. In the case of cross-gamer identification, the observed and tracked gaming records can show the patrons who play both slots and table games. However, the records do not indicate which patrons are likely to become cross-gamers if they currently play slots or table games exclusively.

Another criticism of the RFM analysis is that it focuses only on the three behavioral variables, R, F, and M, in spite of the presence of numerous other variables (e.g., socio-demographic characteristics) and their potential influences on customer behavior (McCarty & Hastak, 2007; Olson et al., 2009). On the other hand, CHAID and logistic regression are not restricted to the RFM variables, and can embrace a variety of variables to predict a specific behavior (Olson et al., 2009).

Gaming Customer Segmentation

With respect to player segmentation, some researchers examined gambling motivations (e.g., Lee, Chung & Bernhard, 2014) while others used casino patrons' transactional gaming data (e.g., Tanford & Suh, 2013) to classify players into different groups. By surveying internet sports gamblers' motivations and passions, Lee at al. identified two distinctive groups with different gambling reasons and outcomes. Players with intrinsic gambling motivations such as excitement, escape and challenge tend to exhibit harmonious gambling passion and consequently experience positive outcomes such as stress release (Lee et al., 2014). On the other hand, players with extrinsic gambling motivations such as money tend to exhibit obsessive gambling passion and therefore experience negative feelings such as guilt and anxiety (Lee et al., 2014). Another study done by Tanford and Suh (2013) examined the gaming volumes associated with different types of casino restaurants (e.g., upscale, casual and buffet) for each of the five different player segments (e.g., high, medium, medium-low, low and untracked) using time-series regression analysis. In their study, customers were grouped into the following five segments based on the player's average daily worth to the subject casino: $400+, $100-$399, $50-$99, $0-$49, and untracked (non-users or non-owners of a player-tracking cards). Furthermore, casino managers review and analyze players' gaming activities, tracked through the casino's information system, in order to decide the amount of complimentary rewards and marketing offers for various segments of the casino's database (Klebanow, 2009). While the segmentation approach based on gaming activities can be easily found in many casinos, some researchers pointed out the potential problems associated with such method (e.g., Iaci & Singh, 2012; Lucas & Kilby, 2008).

Iaci and Singh (2012) noted that clustering or grouping customers by simplistic measures such as zip code or ranked sales can not only ignore a large amount of useful business data but also fail to properly treat the sparse and large dimensional data sets for subsequent analysis (e.g., a data set with 90% zero entries). Similarly, Lucas and Kilby (2008) mentioned that the arbitrary percentage method used in many casinos to assign customers to different groups (e.g., the top 2% of the customer database) fails to consider natural breaks in the player population with reference to t-win. To improve the customer segmentation process, Lucas and Kilby proposed partitioning algorithms that can differentiate the various levels more effectively. Iaci and Singh also introduced mathematical procedures to cluster sparse data sets with a large number of dimensions. Using casino player transaction datasets, Iaci and Singh illustrated how to extract a non- or less-sparse subset from the sparse datasets, which is more suitable for clustering casino patrons.

Regarding the current segmentation practices in the gaming industry, the authors of this article also observed the use of an arbitrary cut-off point at which customer grouping is determined. For example, customers in the entire database are simply sorted by each of the RFM measures and divided into equal groups on each RFM measure. While the arbitrary cut-off points are simple to use, they may not distinguish different groups of customers, and fail to account for the variations of any given variable within a group. It is possible that these variations are greater for that variable than the same between groups. Furthermore, cut-off points that accurately represent the differences between groups could be anywhere. In the CHAID analysis, which is introduced in the following section, the cut-off point is where the greatest degree of differentiation between groups exists and can be established statistically based on the chi-squared test.

Data Mining Approach: CHAID Analysis

Researchers have introduced a variety of statistical and data mining techniques such as CHAID and logistic regression to complement or replace the simple RFM analysis (e.g., Levin & Zahavi, 2001; McCarty & Hastak, 2007; Olson et al., 2009). Among data mining algorithms, decision tree is a popular technique for classification and prediction, and CHAID is one of decision tree algorithms. The CHAID algorithm models differences between groups based on a set of independent variables and identifies variables capable of explaining or predicting a specific behavior (Zuccaro, 2010). In marketing, it is often used to identify customers who are likely to respond to marketing offers and to divide them into segments (Galguera, Luna, & Méndez, 2006; Levin & Zahavi, 2001; McCarty & Hastak, 2007).

CHAID is often described as a three-step process: merging, splitting and stopping (IBM Knowledge Center, 2011a). CHAID develops a model by repeatedly using these three steps on each node, or a subgroup, starting from the root node which contains the entire sample. During the merging stage, CHAID evaluates each pair of categories for a predictor variable to determine whether the pair is significantly different with respect to the dependent variable. With no significant difference, the pair is merged into a single category. In the splitting stage, CHAID evaluates the relationships between the dependent variable and the independent variables, and selects the predictor variable that is most significantly associated with the dependent variable. CHAID uses the Chi-squared ($\chi 2$) statistic to test the statistical significance of the association between the outcome, dependent variable and the predictor variable (SPSS Inc., an IBM Company, 2010).

It divides the original node into subgroups based on the chosen predictor so that the subgroups are mutually exclusive and significantly different with respect to the dependent variable. For numeric (scalar) variables, CHAID identifies the split points, or intervals, at which the greatest degree of separation occurs. The newly formed subgroups, or nodes, are further divided into subsets by another predictor variable that differentiates the target (dependent) variable even further. The splitting procedure stops when no other significant splits exist according to the stopping rules associated with the pre-specified significance level, the limit of a tree depth, and minimum sizes for child and parent nodes. For any group that cannot split further, a prediction is made. Any part of a tree that can no longer split is called a terminal node.

CHAID can handle both categorical and continuous variables (SPSS Inc., an IBM Company, 2010). CHAID displays the modeling results in a graphical tree, making the results easy to interpret. It can create a wider, non-binary classification tree with more than two branches (splits) from a single node (Galguera et al., 2006; SPSS Inc, 2009; Zuccaro, 2010).

Furthermore, classification rules can be generated from the tree, which presents the relationships between the predictor variables and the dependent variable. Because of its relatively simple and easy-to-understand rules, CHAID is often used for customer segmentation and behavior prediction (Galguera et al., 2006; Levin & Zahavi, 2001; McCarty & Hastak, 2007). CHAID is known to be a good fit for the analysis of large data sets (SPSS Inc., 2009).

In the hospitality field, the authors of this study identified only a few studies which used CHAID analysis. One of them is Chung et al. (2004) which illustrated the use of a CHAID model for market segmentation. Using survey data collected from restaurant customers, the authors demonstrated how the CHAID decision tree splits the data into subgroups based on the select predictor variables (e.g., age, income and purpose of visit). The resulting subgroups were significantly different from each other with respect to the dependent variable representing the patronage of hotel restaurants vs. top five casual-dining restaurants.

Applications of Data Mining Methods

In the problem gambling field, researchers employed data mining methods to identify potential problem gamblers (e.g., Braverman, et al., 2013; Dragicevic et al., 2011; Philander, 2013). For example, Braverman et al. used CHAD analysis while Philander compared various data mining methods including logistic regression, artificial neural network, support vector machines and random forest. While data mining methods were applied in the problem gambler detection, to the best knowledge of the authors of this article, they were rarely applied for the purpose of customer classification or behavior prediction in the gaming literature on casino marketing and operations.

Outside the gaming literature, there have been attempts to compare the performances of various data mining methods in predicting a target outcome (e.g., Li et al., 2012; Long et al., 1993). Li et al. built predictive models for diabetic peripheral neuropathy using the clinical data of 274 patients. Of different data mining methods, multilayer perception produced the highest model performance followed by logistic regression (LR) and decision trees (DT) in terms of the Area Under the Receiver Operating Characteristic (ROC) Curves (AUC) which indicates a diagnostic ability of the prediction model in their study. The DT model in their study was constructed by combining the CART and CHAID. Another study by Long et al. also compared the performance of LR to DT in classifying patients with acute cardiac ischemia using the data of 5,773 patients. While the performance of both LR and DT models were comparable to that of the physicians, the LR model outperformed the DT model (Long et al., 1993).

Modeling Cross-Gaming Behavior

Several researchers have examined cross-gaming behavior using the aggregated daily gaming data obtained from hotel-casinos in Nevada and California. Examples of such studies include Ollstein (2006), Lucas (2013) and Suh and Tsai (2012) which examined the effect of poker on slot and/or table gaming volumes. Another study done by Lucas, Dunn, and Kharitonova (2006) examined the effect of bingo on slot gaming volume. Abarbanel (2011) examined the effect of sports- and race-book on slot gaming volume. With a few exceptions, most studies found no significant relationship between the game examined in each of the studies and slot and/or table gaming volume. Furthermore, researchers noted that the positive effects of bingo and poker found in a few cases were relatively small to justify the bingo/poker room operations on the casino floor.

With respect to the cross-gaming behavior between slot and table game players, a recent study done by Suh and Alhaery (2014) appears to be the first study to the best knowledge of the authors of this article. They introduced a model to predict cross-gaming propensity at the individual player level based on the gaming-related behavioral data. Through the logistic regression analysis, the authors found that casino visit frequency and time spent on gaming are closely associated with cross-gaming propensity. Additionally, the greater the amount of money won or lost in gaming the greater the propensity to play both slots and table games. More specifically, cross-gamers in the predominant slot player group were more likely to generate more revenues to both slot and table games. Among predominant table-game players, carnival games and roulette players were more likely to play slots than other table game players.

Regarding the performance of the logistic regression models in correctly classifying cross-gamers and non-cross gamers, Suh and Alhaery (2014) reported that the overall accuracies of the prediction models ranged from 64.92 % to 78.82 %. However, accuracy rates for cross-gamer prediction in their study were lower than those for non-cross-gamer prediction (59.26% vs. 67.57% for the table game player group and 56.38% vs. 80.99% for the slot player group). Additionally, Suh and Alhaery illustrated how to identify the prospects for cross-gaming promotion using the logistic regression models. In conclusion, players who play only slots or table games, but have high-predicted cross-gaming propensities, are the best prospects to target for cross-gaming promotion (Suh & Alhaery, 2014). On the other hand, marketers should avoid marketing to slot- or table-games-only players with low cross-gaming propensities (Suh & Alhaery, 2014).

Research Gap and Purpose of the Study

The lower accuracy rates for predicting cross-gamers relative to those for predicting non-cross-gamers observed in Suh and Alhaery (2014) offer room for improvement. The dearth of research on data mining applications as well as cross-gaming behavior in the gaming literature suggests that more studies need to be undertaken. While data mining analysis might have been performed internally at some casinos or outsourced, data privacy has gained significant importance in the last decade. The paucity of such research could be partially due to the limited access to internal data by individuals outside gaming companies. Hence, this study aimed to test whether a predictive model derived from a data mining technique can accurately predict casino patrons' cross-gaming propensity. If the developed model has an acceptable performance in accurately classifying patrons into one of the two groups: customers playing both slots and table games (cross-gamers) vs. customers playing exclusively slots or table games (non-cross gamers), we also wanted to know the predictors that differentiate these two groups.

Exhaustive CHAID (E-CHAID), which is an enhanced modification of CHAID, was used for modeling cross-gaming behavior in this study. E-CHAID performs a more thorough analysis and segmentation by examining all possible splits for each predictor that maximizes the final model accuracy, and thus often requires a longer computing time to build a tree than CHAID (IBM Knowledge Center, 2012). For comparison purposes, the same data and variables used in Suh and Alhaery (2014) for logistic regression analysis in modeling cross-gaming behavior were analyzed in this study. If the E-CHAID model performs better than the logistic regression model in terms of prediction accuracy, the E-CHAID model can then be an alternative to the logistic regression model in predicting cross-gaming propensity.

While both E-CHAID and logistic regression methods can examine the relationship between the dependent, or target, variable and a set of independent variables, there are advantages and disadvantages of using each method. CHAID is known to be well suited for dealing with missing data as it treats a missing value as a separate category in estimating the model (SPSS Inc., an IBM Company, 2010). Furthermore, CHAID is suitable for large data sets. Logistic regression, on the other hand, works well with smaller data sets and generates robust parameter estimates (Zuccaro, 2010). However, logistic regression tends to require a higher level of data integrity to make predictions. For example, log transformation is often performed on the predictor variable with a highly skewed distribution in order to reduce the skewness. Outliers can also disproportionally affect the performance of the logistic regression model.

Methods

Data

The same data set used in Suh and Alhaery (2014) was used for the present study. In their study, the property which donated the data was described as a hotel-casino located on the Las Vegas Strip, Nevada, offering a variety of resort amenities mainly to tourists. The property generates substantial revenues from its non-gaming amenities, while slot machines are the main source of gaming revenues. The data set which consists of 14,120 players who were randomly selected from the property's customer database included individual players' gaming records in the year of 2008, such as the amount of money won or lost in gaming, casino trip frequency, the amount of time spent on gaming, and theoretical gaming win (t-win) during the 12-month sample period. For analysis, the players in the data set were separated into two groups by the type of game (slot or table) that a player primarily played during the sample period. This is the same procedure performed in Suh and Alhaery (2014), and it was followed in this study for comparison purposes. This simple division is also what the authors of this article observed in the gaming industry. For example, a player with a higher average t-win per trip for slots than for table games during the sample period was categorized as a Predominant Slot Player. Otherwise, the patron was classified as a Predominant Table Game Player. There were a few cases where the t-win values for table games and slots were identical, and they were excluded from analysis. As a result, each of the two groups contained patrons who played both table games and slots as well as those who played only slots or table games. Table 1 shows the distribution of the original data set. As shown, the slot player set contained 10,651 players while the table-game player set consisted of 3,469 players. Given that 62.5% - 90.3% of gaming revenues are from slots across different states in the U.S. (American Gaming Association, 2013), the distribution of players between slots and table games at the subject property appears to reflect the revenue pattern in the U.S. gaming market.

Table 1
Data Distribution

Sample size (%) by Category	Predominant Slot Players		Predominant Table Game Players	
	Original	Balanced	Original	Balanced
Cross-gamers	940 (9%)	9,711 (50%)	1,107 (32%)	2,362 (50%)
Non-cross-gamers	9,711 (91%)	9,706 (50%)	2,362 (68%)	2,364 (50%)
Total	10,651 (100%)	19,417 (100%)	3,469 (100%)	4,726 (100%)

Sample Balancing

The proportion of cross-gamers was much smaller than that of non-cross gamers in both slot and table game player groups. As shown in Table 1, 32% (1,107) of the 3,469 predominant table game players played both slots and table games during the sample period. In the predominant slot group, only 9% (940) played both slots and table games. Because of the observed imbalance between the cross-gamer and non-cross-gamer samples, the sample size in each group was balanced by duplicating the existing records in a smaller category, which in this case is the cross-gamer group. With balancing, sample sizes became approximately equal for each group. Table 1 shows the results of sample balancing in the Balanced columns. Sample balancing is often performed in order to better capture the events that occur less frequently. This is because a predictive model developed based on the unbalanced data tends to classify cases into the category that occurs more frequently in order to achieve higher prediction accuracy (SPSS Inc., 2009). However, overall accuracy is less critical in this study because the goal for this study is to predict and identify the customers with the highest cross-gaming propensity. Furthermore, casino marketers would be interested in the same outcome considering the potential gaming revenue from cross-gaming prospects vs. lower direct-mail costs. While the predictive model was developed based on the balanced sample, the evaluation of the model in terms of its classification accuracy was performed based on the original, unbalanced sample.

Variables

The dependent variable was a dichotomous variable indicating the absence or presence of cross-gaming activity for each player during the sample period. For example, the dependent variable in the predominant slot player group was named as BIorNoBI to indicate the presence of buy-in for a given slot player. Buy-in indicates the amount of money spent on purchasing gaming chips for table game play. If a slot player has any record of buy-in during the sample period, the player was coded as a '1' meaning cross-gamer. All other players were coded with a '0', signifying non-cross-gamers. For the predominant table game players, the dependent variable was named as CIorNoCI, to indicate the presence of coin-in for a given table game player. Coin-in indicates the amount of money wagered on slots. If a table game player has any record of coin-in during the sample period, the player was coded as one (cross-gamer). If not, that played was coded with a zero (non-cross-gamer). Table 2 presents a brief description of the independent variables used for the E-CHAID analysis and the matching concepts tested in Suh and Alhaery (2014). Variable names are displayed in parenthesis.

Table 2
Variable Used in Analysis

Concept	Variable Name & Description
Trip frequency	The number of casino trips during the sample period (Trips_within_12) and the same for the entire period (lifetime) that a player patronized the subject casino (LTD_Trips)
Length of the player-casino relationship	The number of days since the player's first trip to the subject casino (Days_since_first)
Recency of a casino trip	The number of days since the player's most recent trip to the subject casino (Days_since_last)
Dollar amount of gaming volume for slots	The average dollar amount wagered on slots, or coin-in, per trip during the 12-month sample period (CI12pertrip) and the same for the lifetime (CILTDpertrip)
Dollar amount of gaming volume for table games	The average dollar amount used to purchase gaming chips, or buy-in, at table games per trip during the sample period (BI12pertrip) and the same for the lifetime (BILTDpertrip)
Dollar amount of theoretical win for table games	The average t-win per trip for table games during the 12-month sample period (Ttheo12pertrip) and the same for the lifetime (TtheoLTDpertrip)
Dollar amount of theoretical win for slots	The average t-win per trip for slots during the 12-month sample period (Stheo12trip) and the same for the lifetime (StheoLTDpertrip)
Table game type	The predominant table game that a patron mainly played during the sample period (Blackjack, Craps, Baccarat, Roulette, Carnival Games)
Time spent on gaming	The average amount of time in minutes that a player spent on gaming per trip during the sample period (Min12pertrip)
Dollar amount won or lost in gaming	The average dollar amount that a player won or lost in gaming per trip during the sample period (Act12pertrip)
Play speed	The dollar amount that a player won or lost per minute during the 12-month sample period (Act_within_12_PER_MIN)

Analysis

The E-CHAID algorithm was applied to the gaming data using the IBM SPSS Modeler 15.0. In the E-CHAID model, the predicted membership of a player, whether the player is a cross-gamer or not, was determined based on a set of independent variables. The variable that was most significantly associated with the dependent variable was chosen as the best and first predictor to split the sample population. Each of the subgroups was further evaluated for a possible split based on another predictor variable. If there were significance differences in cross-gaming activity based on the Chi-Squared test statistic and the size of the subpopulation, the subgroup was further divided into smaller groups according to the selected predictor variable. This process was repeated for each subset until no further splits were statistically significant. Modeling procedures were performed at the default settings in which the significance level for splitting and merging was the 0.05 alpha level, and the minimum number of cases in parent and child nodes were set to 2.0% and 1.0% of the sample, respectively. The depth of E-CHAID trees was set to a maximum of 5 levels to produce a parsimonious model while avoiding over-fitting. At the end of splitting, E-CHAID produced a classification tree which consisted of a series of groups that were different from one another with respect to the cross-gaming propensity. Additionally, the derived prediction model generates the cross-gaming propensity scores for each patron in the sample. Propensity scores which range from zero to one indicate the probability of a given player to cross-play. The propensity score close to zero indicates the low probability of cross-gaming while the propensity score close to one indicates the high probability of cross-gaming.

Results

Cross-Gaming Behavior

Table 3 presents the means (M) and standard deviations (SD) of cross-gamers vs. non-cross-gamers in each of the groups of predominant slot and table game players. SD is shown in parentheses. For the variables representing table game types (e.g., Blackjack), frequencies and percentages of different table game players are displayed instead of M and SD. A comparison between cross-gamers and non-cross-gamers revealed that cross-gamers made more recent and frequent trips to the subject casino and spent more time on gaming than non-cross-gamers. See the section on descriptive statistics in Suh and Alhaery (2014) for more details.

Table 3
Gaming Behaviors of Table Game-Only, Slot-Only, and Cross-Gamers

Variables	Predominant Table Game Players		Predominant Slot Players	
	Table-Only	Cross-Gaming	Slot-Only	Cross-Gaming
LTD_Trips	2.3 (3.1)	3.6 (9.4)	3.4 (11.1)	5.2 (16.4)
Trips_within_12	1.3 (1.0)	1.7 (2.6)	1.8 (4.9)	2.5 (6.9)
BILTDpertrip	3,571.2 (15,291.7)	3,241.2 (13,248.3)	--	--
CILTDpertrip	--	--	1,509.9 (11,751.9)	5,597.4 (24,113.4)
BI12pertrip	4,078.0 (20,626.3)	3,578.7 (15,144.4)	--	--
CI12pertrip	--	--	1,426.1 (12,004.6)	6,272.9 (27,103.2)
TtheoLTDpertrip	686.6 (3,636.4)	614.0 (2,919.0)	--	--
StheoLTDpertrip	--	--	103.7 (471.7)	347.4 (1,130.2)
Ttheo12pertrip	787.5 (4,918.4)	583.6 (2,932.5)	--	70.0 (208.6)
Stheo12pertrip	--	79.6 (732.2)	97.2 (415.8)	387.0 (1,289.6)
Days_since_first	434.4 (418.1)	481.4 (443.2)	411.7 (407.9)	514.2 (445.9)
Days_since_last	162.1 (110.7)	142.9 (105.1)	160.3 (107.3)	138.0 (104.7)
Min12pertrip	192.0 (267.7)	311.3 (360.3)	105.8 (186.3)	316.5 (408.2)
Act12pertrip	731.0 (12,166.9)	915.0 (6,810.9)	89.7 (782.0)	560.4 (2,308.8)
Act_within_12_PER_MIN	7.0 (112.2)	3.2 (29.4)	0.7 (8.6)	1.7 (5.3)
Craps	284 (71%)	118 (29%)	--	--
Roulette	267 (60%)	179 (40%)	--	--
Blackjack	1,371 (73%)	506 (27%)	--	--
Baccarat	98 (76%)	31(24%)	--	--
Carnival games	342 (56%)	273 (44%)	--	--

Slot Gaming Prediction

The E-CHAID model for predicting slot gaming among predominant table game players produced both a graphical tree and classification rules. The resultant E-CHAID tree displayed 46 terminal nodes. A review of the tree revealed that ten of 16 input variables included in the E-CHAID analysis were statistically significant and used for building a cross-gaming prediction model. The first predictor chosen by the E-CHAID model in separating cross-gamers from non-cross gamers was the Min12pertrip variable ($\chi 2$ = 379.753, df = 5, p = 0.000). The subgroups, or nodes, were further divided into subsets according to the Act_within_12_PER_MIN, Trips_within_12, and BILTDpertrip variables. The E-CHAID algorithm derived from IBM SPSS Modeler produces the relative importance of each predictor in estimating the model. Table 4 presents the importance order of significant variables to differentiate between cross-gamers and non-cross gamers.

Table 4
Predictor Importance in Cross-Gamer Classification

Importance Order	Slot Gaming Prediction	Table Gaming Prediction
1	Min12pertrip	Act12pertrip
2	BILTDpertrip	CI12pertrip
3	Act_within_12_PER_MIN	Stheo12trip
4	Trips_within_12	Min12pertrip
5	TtheoLTDpertrip	Trips_within_12
6	BI12pertrip	StheoLTDpertrip
7	Blackjack	Act_within_12_PER_MIN
8	Carnival games	LTD_Trips
9	Roulette	CILTDpertrip
10	LTD_Trips	Days_since_first
11	--	Days_Since_last

Classification rules can be generated by following the paths or branches from the root of the E-CHAID tree. As shown in Figure 1, the first split is on Min12pertrip. A prediction is made in the terminal node where a branch of the tree ends. Since there are 46 terminal nodes in the tree, 46 classification rules can be derived from the model. Table 5 displays sample rules along with the record count and percentage accuracy for each rule. For example, the classification rules for one of the terminal nodes in the third line of the decision tree, Node 71, are listed in Table 5 (See Rule 21 for 1.0). Rule 21 for 1.0 shows that of the 137 patrons in this node, there are 105 individuals exhibiting gaming minutes per trip for the 12-month sample period greater than 378, the average actual win or loss per minute for the sample period greater than -5.594 and less than or equal to 12.488; playing Blackjack; and having the average t-win per trip for table games during the 12-month sample period less than or equal to 204. These 105 players were predicted by the E-CHAID model to be cross-gamers. The confidence or accuracy rate of this prediction is shown in parentheses, indicating the percentage of players who were correctly classified as cross-gamers (0.766 = 105/137).

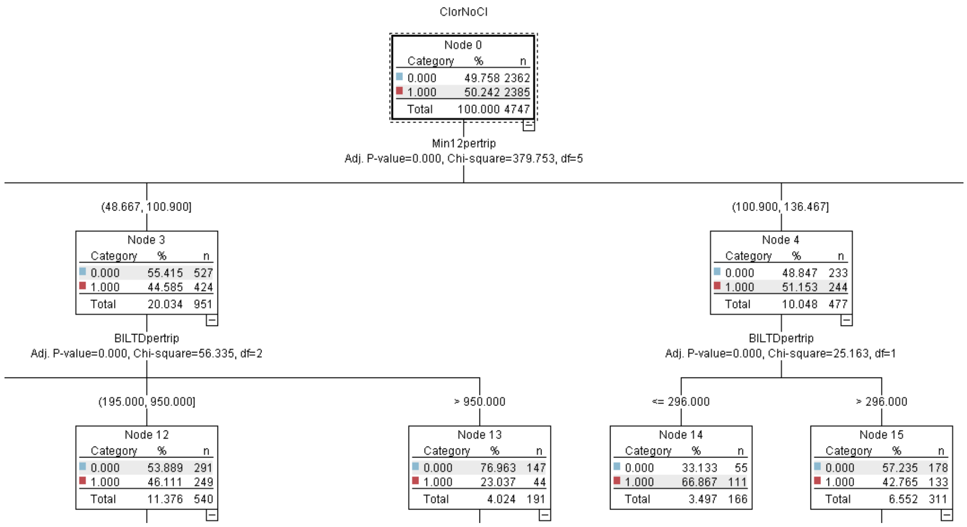

Figure 1
E-CHAID Decision Tree for Classifying Cross-Gamers in the Table-Game Player Group (Excerpt)

Table 5
E-CHAID Classification Rules (excerpt)

Slot Play Prediction	Table Game Play Prediction
Rules for 0 - contains 23 rule(s) Rule 1 for 0.0 (80; 0.688) if Min12pertrip <= 28 and Trips_within_12 <= 1 and BILTDpertrip <= 100 and Blackjack = 0 then 0.000 Rule 2 for 0.0 (78; 0.846) if Min12pertrip <= 28 and Trips_within_12 <= 1 and BILTDpertrip <= 100 and BJ = 1 then 0.000 ~ ~ ~ ~ Rule 23 for 0.0 (65; 0.615) if Min12pertrip > 378 and Act_within_12_PER_MIN > 12.488 then 0.000 Rules for 1 - contains 23 rule(s) Rule 1 for 1.0 (72; 0.528) if Min12pertrip <= 28 and Trips_within_12 > 1 then 1.000 Rule 2 for 1.0 (98; 0.643) if Min12pertrip > 28 and Min12pertrip <= 48.667 and Trips_within_12 > 1 then 1.000 ~ ~ ~ ~ Rule 21 for 1.0 (137; 0.766) if Min12pertrip > 378 and Act_within_12_PER_MIN > -5.594 and Act_within_12_PER_MIN <= 12.488 and Blackjack = 1 and TtheoLTDpertrip <= 204 then 1.000 ~ ~ ~ ~ Rule 23 for 1.0 (149; 0.664) if Min12pertrip > 378 and Act_within_12_PER_MIN > -5.594 and Act_within_12_PER_MIN <= 12.488 and Blackjack = 1 and TtheoLTDpertrip > 204 and Min12pertrip > 630.250 then 1.000	Rules for 0 - contains 31 rule(s) Rule 1 for 0.0 (272; 0.581) if Act12pertrip <= -102 and Min12pertrip <= 46.617 then 0.000 Rule 2 for 0.0 (251; 1.0) if Act12pertrip > -102 and Act12pertrip <= -11 and ciltdpertrip <= 45 and Min12pertrip <= 14.517 then 0.000 ~ ~ ~ ~ Rule 31 for 0.0 (202; 0.693) if Act12pertrip > 109.500 and Act12pertrip <= 201.500 and ciltdpertrip > 967 and Trips_within_12 <= 1 and Act_within_12_PER_MIN > 0.695 then 0.000 Rules for 1 - contains 34 rule(s) Rule 1 for 1.0 (304; 0.641) if Act12pertrip <= -102 and Min12pertrip > 46.617 and Min12pertrip <= 135.483 and LTD_Trips <= 1 then 1.000 Rule 2 for 1.0 (318; 0.78) if Act12pertrip <= -102 and Min12pertrip > 46.617 and Min12pertrip <= 135.483 and LTD_Trips > 1 then 1.000 Rule 3 for 1.0 (424; 0.825) if Act12pertrip <= -102 and Min12pertrip > 135.483 and Min12pertrip <= 540.167 and Act_within_12_PER_MIN <= - 1.438 then 1.000 ~ ~ ~ ~ Rule 34 for 1.0 (396; 0.934) if Act12pertrip > 853 and Min12pertrip > 300.125 and stheoltdpertrip > 448.667 and Trips_within_12 > 1 then 1.000

Classification Accuracy for Slot Gaming Prediction

The performance of the E-CHAID model can be measured by comparing the predicted membership of a subject to the actual, observed group membership (SPSS Inc, 2009). The E-CHAID model for slot player prediction produced the overall accuracy rate of 69.21% by correctly classifying 2,401 of a population of 3,469 players. The proportion of the table game players who were correctly predicted as non-cross gamers was 70.79% (1,672 of 2,362 table game only players) while the proportion of the cross-gamers who were correctly predicted to play both slots and table games was 65.85% (729 of 1,107 observed cross-gamers). Additionally, the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) was used to assess the overall performance of the model in terms of its prediction accuracy. The AUC value of the model was 0.748 ($p = 0.000$) indicating that the model has good predictive power.

Table Gaming Prediction

The E-CHAID model for predicting table gaming among predominant slot players produced 65 terminal nodes. A review of the E-CHAID tree revealed that the first predictor chosen by the E-CHAID model in separating cross-gamers from non-cross gamers was Act12pertrip ($\chi2 = 5065.260$, $df = 9$ $p = 0.000$). The subgroups, or nodes, were further divided into subsets based on other predictors such as the CI12pertrip, Min12pertrip, and Stheo12trip variables. Table 4 presents the importance order of significant variables to differentiate between cross-gamers and non-cross gamers for the table-gaming prediction model.

With respect to the E-CHAID tree, the tree generated 65 terminal nodes, hence 65 classification rules. The first split was made on the Act12pertrip variable. Table 5 shows sample classification rules. For example, the classification rules for Node 48 which is one of the terminal nodes in the fourth line of the tree is shown in Rule 2 for 1.0 in Table 5. The rule shows that if Act12pertrip <= -102, 46.617 < Min12pertrip <= 135.483, and LTD_Trips > 1, the model predicts a patron as most likely a cross-gamer. In other words, if the average actual amount of win or loss per trip during the 12-month sample period is less than or equal to -102; the average gaming minutes per trip during the sample period is greater than 46.617 and less than or equal to 135.483; and the number of casino trips during a player's lifetime is greater than 1, then the model value for BIorNoBI is 1, meaning a table game player who also plays slots. The rule further reveals that of the 318 cross-gamers in this node, 248 individuals were correctly predicted as cross-gamers (78% accuracy).

Classification Accuracy for Table-Gaming Prediction

The E-CHAID model for predicting table gaming among predominant slot players produced an overall accuracy rate of 71.91%. The model correctly classified 7,659 of 10,651 players. The proportion of the slot only players who were correctly classified as non-cross gamers was 70.81% (6,876 of 9,711 slot only players) while the proportion of cross-gamers who were correctly predicted to play both slot and table games was 83.30% (783 of 940 cross-gamers). As stated earlier, the model's classification accuracy is based on the unbalanced, original data. The AUC value for the model was 0.853 ($p = 0.000$).

Table 6
Classification Accuracy Comparison between E-CHAID and Logistic Regression Models

	Slot Gaming Prediction		Table Gaming Prediction	
Category (%)	E-CHAID	Logistic Regression	E-CHAID	Logistic Regression
Non-Cross-Gamer	70.79	67.57	70.81	80.99
Cross-Gamer	65.85	59.26	83.30	56.38
Overall	69.21	64.92	71.91	78.82

Discussion

The E-CHAID models for cross-gamer prediction produced overall accuracies that are acceptable along with satisfactory AUC values. Based on these results, it can be concluded that cross-gaming propensity can be predicted by the E-CHAID model using gaming-related behavioral data. Furthermore, a comparison of the classification accuracies between the E-CHAID models in this study and the logistic regression models in Suh and Alhaery (2014) indicated that E-CHAID outperformed logistic regression in classifying cross-gamers accurately. Table 6 presents the accuracy comparison between the two studies.

As shown, the E-CHAID model had a higher overall accuracy rate (69.21%) than the logistic regression model (64.92%) in predicting slot play propensity among predominant table game players. The E-CHAID model also outperformed the logistic regression model in both categories of predicting cross- and non-cross-gamers. The accuracy rate for the cross-gamer classification was higher in the E-CHAID model than in the logistic regression model (65.85% vs. 59.26%). The same was observed in the non-cross gamer group (70.79% vs. 67.57%). In predicting table-gaming propensity among predominant slot players, the logistic regression model had a higher overall accuracy rate (78.82%) than the E-CHAID model (71.91%). This was mainly because the logistic regression model outperformed the E-CHAID model (80.99% vs. 70.81%) in predicting non-cross gamers, or slot-only players. As the size of the non-cross gamer group was much bigger than that of the cross-gamer group, the classification accuracy rate in the larger group (non-cross-gamer group) rather than in the smaller group (cross-gamer group) had a greater effect on the overall accuracy. In other words, high accuracy rates in the non-cross-gamer group resulted in an increase in overall accuracy rates. However, the E-CHAID model outperformed the logistic regression model in predicting cross-gamers, (83.30% vs. 56.38%).

In summary, it can be concluded that the E-CHAID analysis can deliver higher accuracy in predicting cross-gaming prospects than a conventional modeling technique, logistic regression analysis. Finally, comparisons of the model results between slot and table gaming prediction models and between E-CHAID and logistic regression models revealed that the following variables appear consistently across all models for cross-gaming prediction: The average amount of time spent on gaming per trip during the 12-month sample period (Min12pertrip), the dollar amount won or lost per minute during the sample period (Act_within_12_PER_MIN), the total number of casino trips since the player patronized the subject casino (LTD_Trips), and the number of casino trips during the sample period (Trips_within_12).

Managerial Implications

Cross-Gaming Propensity Scores

In addition to the predicted membership, the E-CHAID model generated cross-gaming propensity scores for each patron. A propensity score represents the estimated probability of a subject being a cross-gamer. A high propensity score for any given patron in this study means a high likelihood of cross-gaming for that individual. This information is crucial given that additional gaming revenues can be generated by cross-gamers who can be converted from slot or table game only players. Hence, casino managers can use the propensity scores derived from the E-CHAID model to identify prospects with high chances of cross-gaming and convert them into cross-gamers with an appropriate offer.

Target Marketing for Cross-Gaming

Casino managers should pay attention to the players who are not currently cross-gamers but predicted to be cross-gamers, especially the ones with high cross-gaming propensities. While these patrons currently play slots or table games only, these prospects have a higher chance to respond to a cross-gaming promotion and to be converted to become cross-gamers. Converting them into cross-gamers can possibly increase the proportion of cross-gamers within the population and eventually gaming revenues if the converted cross-gamers spend more on playing different games. In fact, the descriptive statistics comparing cross-gamers and non-cross gamers show more frequent and recent visits among cross-gamers as well as larger spending on gaming especially among predominant slot players who cross-play table games. On the contrary, converting exclusive slot or table game players, especially the ones with low cross-gaming propensities would be very difficult and costly to the casino.

In summary, casino marketers can identify cross-gaming prospects based on patron's predicted memberships and propensity scores derived from the E-CHAID models. This, in turn, will help them develop customized marketing strategies for the targeted prospects and make more efficient use of marketing dollars. Sending cross-gaming offers to all patrons regardless of their cross-gaming propensities would result in high costs and low redemption rates. By selectively targeting patrons with high cross-gaming propensities, however, marketers could achieve higher response rates and returns on marketing investment.

Estimating Cross-Gaming Revenues

Once cross-gaming prospects are identified, marketing managers need to determine the costs for targeting the prospects as well as the expected incremental revenues from cross-gaming. Evaluating economic benefits from cross-gaming is crucial to allocate a casino's limited resources more efficiently and to create more effective marketing programs. One way to estimate incremental gaming revenues from cross-gaming prospects is to multiply the average cross-gaming revenue per trip by the number of cross-gaming prospects. For example, the incremental slot revenues from the cross-gamers in the predominant table-game player group can be estimated by multiplying the average slot revenue per trip per cross-gamer in the group by the number of cross-gaming prospects. Table 1 presents the average slot t-win per trip (Stheo12trip) per predominant table game player during the sample period is $79.6. The number of cross-gaming prospects in the predominant table-game player group can be determined based on the cross-gaming propensity scores generated by the E-CHAID model.

The same calculation can be performed to approximate the incremental table gaming revenue from the cross-gaming prospects in the predominant slot-player group. For example, if the marketing manager at the subject property decided to target 30% of the 10,651 predominant slot players, and 10% of the target would be converted to cross-gamers, a marketing manager can expect to generate the incremental table gaming revenue of $22,367 from cross-gamers during their single trips. The following example illustrates the calculation: $70 of the average table game t-win per trip per slot player (Ttheo12pertrip in Table 1) x 320 (30% of 10,651 x 10%). While our example is limited to the sample of slot players used in this study, the impact of cross-gaming on the bottom line could be much greater with a bigger customer database and a higher conversion rate through targeted marketing.

The above-mentioned calculations, however, do not take the substitution of gaming budgets into consideration. In other words, a newly converted cross-gamer who used to play only slots may substitute his or her slot gaming budget for table gaming instead of increasing his or her overall gaming budget. Other new cross-gamers may also split their gaming budgets between slots and table games instead of increasing their overall gaming budgets. Hence, such substitution or split of gaming budgets could affect the incremental gaming contribution. Nevertheless, the overall increase in gaming revenue through increased cross-gaming play is still a possibility for the subject property. This is mainly because the combined t-win (table game t-win + slot t-win) of cross-gamers is comparable or higher than the t-win of non-cross-gamers especially in the predominant slot player group. As shown in Table 3, the combined t-win of the cross-gamers during the 12-month sample period in the predominant slot player group was $457.0 (387.0 + 70.0) while the t-win of the slot-only players was $97.2. With respect to the predominant table game player group, the t-win of the table only players was $787.5 while the combined t-win of the cross-gamers was $663.2 (583.6 + 79.6).

Limitations and Suggestions for Future Research

This study is not without limitations. The prediction models advanced herein should be applied to other data sets and casinos. While the predictors identified in this study could be useful in predicting cross-gaming behavior, future researchers should include other variables in the model. For example, socio-demographic data could reveal other predictors and improve the model's prediction accuracy. Additionally, other modeling methods can be explored for cross-gaming prediction, and their performances can be compared. This, in return, would provide a better understanding of cross-gaming behavior and help researchers and casino managers develop classification models with higher prediction accuracy.

References

Abarbanel, B.L.L., Lucas, A.F., & Singh, A.K. (2011). Estimating the indirect effect of sports books on other in-house gaming volumes. *UNLV Gaming Research & Review Journal*, *15*(2), 77-90.

American Gaming Association (2013). *2013 Sate of the states: The AGA survey of casino entertainment*. Retrieved from http://www.americangaming.org/sites/default/files/uploads/docs/aga_sos2013_fnl.pdf

Brokopp, J. (2008). *Will new video poker game attract slot players?* Retrieved Gaming Guru: http://brokopp.casinocitytimes.com/article/will-new-video-poker-game-attract-slot-players?-40694

Braverman, J., LaPlante, D. A., Nelson, S. E., & Shaffer, H. J. (2013). Using cross-game behavioral markers for early identification of high-risk internet gamblers. *Psychology of Addictive Behaviors, 27*(3), 868-877. doi: 10.1037/a0032818

Burton, E. (2008). bwin And Harvard Sign Research Cooperation Agreement. Retrieved from http://www.pokernewsdaily.com/bwin-and-harvard-sign-research-cooperation-agreement-830/

Casino Enterprise Management. (2010). *FutureLogic—TableXchange®.* Retrieved from http://www.casinodataimaging.com/articles/ArticleName22.pdf

Chung, H. M., & Gray, P. (1999). Special section: Data mining. *Journal of Management Information Systems*, 11-16.

Dragicevic, S., Tsogas, G., & Kudic, A. (2011). Analysis of casino online gambling data in relation to behavioural risk markers for high-risk gambling and player protection. *International Gambling Studies*, *11*(3), 377-391. doi: http://dx.doi.org.ezproxy.fiu.edu/10.1080/14459795.2011.629204

Fleming, A. (2006). *Online and mobile gambling market momentum*. Retrieved from InvestorIdeas.com: http://www.investorideas.com/Articles/Gambling_Market_Momentum.asp

FutureLogic, Inc. (2012). *FutureLogic's TableXchange printer/scanner wins award at G2E*. Retrieved from http://www.futurelogicinc.com/newsAndEvents/futureLogicNews/detail.aspx?newsID=244

Experfy Insights. (2014). *Big data and the transformation of the gaming industry*. Retrieved from http://www.experfy.com/blog/big-data-transformation-gaming-industry/

Galguera, L., Luna, D., & Méndez, M. P. (2006). CHAID to segment loyalty card holders. *International Journal of Market Research , 48*(4), 459-479.

Green, M. (2010). *Games: Beyond the green felt. Interblock's organic blackjack*. Retrieved from Casino Journal: http://www.slotmanagermag.com/Articles/Feature_Articles/BNP_GUID_9-5 2006_A_10000000000000941657
Iaci, R., & Singh, A. K. (2012). Clustering high dimensional sparse casino player tracking datasets. *UNLV Gaming Research & Review Journal, 16*(1), 21-43.

IBM Knowledge Center. (2011a). *CHAID and exhaustive CHAID algorithms*. Retrieved from http://pic.dhe.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=/com.ibm.spss.statistics.help/alg_tree-chaid.htm
IBM Knowledge Center. (2011b). *RFM analysis*. Retrieved from IBM Corporation: http://pic.dhe.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=/com.ibm.spss.statistics.help/rfm_intro.xml.htm
IBM Knowledge Center. (2012). *CHAID node*. Retrieved from http://www-01.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.modeler.help/chaidnode_general.htm
IGT. (2011). *Triple Play™ Five Play® Wheel Poker®.* Retrieved from IGT: http://media.igt.com/marketing/PromotionalLiterature/GamePromoLit_11438-18731.pdf
Klebanow, A. (2009). Customer reinvestment analysis in the gaming industry. *Worldwide Hospitality and Tourism Themes, 1*(4), 386-391. doi: http://dx.doi.org/10.1108/17554210911002228
Lee, C., Chung, N., & Bernhard, B. J. (2014). Examining the structural relationships among gambling motivation, passion, and consequences of internet sports betting, *Journal of Gambling Studies, 30*(4), 845-858. doi: 10.1007/s10899-013-9400-y
Levin, N., & Zahavi, J. (2001). Predictive modeling using segmentation. *Journal of Interactive Marketing, 15*(2), 2-22.
Lucas, A. F. (2004). Estimating the impact of match-play promotional offers on the blackjack business volume of a Las Vegas Hotel Casino, *Journal of Travel & Tourism Marketing*, 17 (4), 23-33.
Lucas, A. F. (2013). Estimating the impact of Las Vegas poker rooms on slot and table game business volumes: An empirical examination of extant operating theory. *Cornell Hospitality Quarterly, 52*(2), 209-218. doi:10.1177/1938965513483921
Lucas, A. F., & Kilby, J. (2008). *Principles of casino marketing.* Norman, OK: Okie International.
Lucas, A. F., Dunn, W. T., & Kharitonova, A. (2006). Estimating the indirect gaming contribution of bingo rooms. *UNLV Gaming Research & Review Journal, 10*(2), 39-54.
McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of business research , 60*(6), 656-662.
Mutyala, S. (2011). *Using RFM to identify your best customers*. Retrieved from Eight Leaves: http://www.eightleaves.com/2011/01/using-rfm-to-identify-your-best-customers
Ollstein, B. W. (2006). Estimating the indirect gaming contribution of poker rooms: Is poker room volume a peripheral driver of slot revenue?, MS thesis, University of Nevada, Las Vegas.
Olson, D. L., Cao, Q., Gu, C., & Lee, D. (2009). Comparison of customer response models. *Service Business, 3*(2), 117-130.
Peacock, P. R. (1998). Data mining in marketing: Part 1. *Marketing Management, 6*(4), 8-18.
Philander, K. S. (2014). Identifying high-risk online gamblers: A comparison of data mining procedures, *International Gambling Studies*, *14*(1), 53-63, doi: 10.1080/14459795.2013.841721
Pollard, J. (2008). *Party gaming maximises on cross-selling with deals on movie-themed slots*. Retrieved from http://www.egrmagazine.com/news/12991/partygaming-maxi-

mises-on-crossselling-with-deals-on-moviethemed-slots.thtml

SAS. (2010). *Independent research firm recognizes SAS as a leader in predictive analytics, data mining*. Retrieved from SAS: http://www.sas.com/offices/asiapacific/sp/news/releases/forresterwave-predictiveanalytics-datamining.html

SPSS Inc. (2009). *Predictive Modeling with PASW® Modeler.* Chicago, IL.: SPSS Inc.

SPSS Inc., an IBM Company. (2010). *Introduction to PASW® Modeler and data mining*. Retrieved from Chicago, IL.

Suh, E., & Alhaery, M. (2014). Maximizing player value through the application of cross-gaming predictive models. *International Journal of Contemporary Hospitality Management, 26*(8), 1243-1269.

Suh, E. and Tsai, H. (2012). Examining the relationship between poker and the gaming volumes of Las Vegas Hotel Casinos. *Journal of Business Research*. 66 (2013), 1651-1658. doi: 10.1016/j.jbusres.2012.12.011

Tanford, S. & Suh, E. (2013). How restaurant variety indirectly impacts gaming for different casino worth segments, *International Journal of Contemporary Hospitality Management*, *25*(3), 328-345. doi: http://dx.*doi*.org/10.1108/09596111311311008.

Woodie, A. (2011). *Las Vegas Sands taps SAS for predictive analytics*. Retrieved from Four Hundred Stuff: http://www.itjungle.com/fhs/fhs090611-story07.html

Zuccaro, C. (2010). Classification and prediction in customer scoring. *Journal of Modelling in Management , 5*(1), 38-53.