Extended RFM logit model for churn prediction in the mobile gaming market

Ana Perišić[1,*] **and Marko Pahor**[2]

[1] *Polytechnic of Šibenik, 22000 Šibenik, Croatia*
E-mail: ⟨*sisak@vus.hr*⟩

[2] *University of Ljubljana, School of Economics and Business*
10000 Ljubljana, Slovenia
E-mail: ⟨*marko.pahor@ef.uni-lj.si*⟩

Abstract. As markets are becoming increasingly saturated, many businesses are shifting their focus to customer retention. In their need to understand and predict future customer behavior, businesses across sectors are adopting data-driven business intelligence to deal with churn prediction. A good example of this approach to retention management is the mobile game industry. This business sector usually relies on a considerable amount of behavioral telemetry data that allows them to understand how users interact with games. This high-resolution information enables game companies to develop and adopt accurate models for detecting customers with a high attrition propensity. This paper focuses on building a churn prediction model for the mobile gaming market by utilizing logistic regression analysis in the extended recency, frequency and monetary (RFM) framework. The model relies on a large set of raw telemetry data that was transformed into interpretable game-independent features. Robust statistical measures and dominance analysis were applied in order to assess feature importance. Established features are used to develop a logistic model for churn prediction and to classify potential churners in a population of users, regardless of their lifetime.

Keywords: churn prediction, logistic regression, mobile games, RFM

---

1. Introduction

Free-to play (F2P) games and other service-based business models are dominant in the online gaming market today. Within the F2P business setting, instead of charging a single upfront fee for the game license, companies make money with subscriptions, advertising, or microtransactions from paying players throughout their lifetime as customers [6]. F2P games are free to download and the developer hopes to generate revenue via in-app purchases and/or advertising [20]. For this reason, F2P model only works if the first taste that users get is so compelling that they regularly come back for more [11]. Thus, customer retention is crucial for successful monetization of a game. Predicting player behavior is one of the central and most common challenges in game analytics. This is notably due to the mobile F2P section of the game industry being to a degree dependent on predictive modelling to be able to monitor, control and forecast their revenue flow [27].

In user-oriented game analytics, the most important data source is telemetry. Telemetry is data, typically digital, obtained over a distance, and in a common scenario of digital games,

---

*Corresponding author.

an installed game client is transmitting data about user-game interaction to a collection server, where the data is transformed and stored in a format that supports rapid analysis and reporting [10]. Mobile applications and, especially, mobile games are highly integrated into humans' lives. They are used in different environment, such as home, work, restaurants, offices, public transport, or "on the go". For this reason, telemetry is especially beneficial since it provides the close observation of users in everyday context. Moreover, behavioral telemetry offers high-resolution, precise data about how users interact with games and the application of telemetry to obtain data about in-game and associated behavior from game players has gained substitutional traction in both the industry and academic sector [9]. Telemetry datasets allow for recording any user action of interest and capture a full range of temporal characteristics of user behavior. For example, events such as "app open", "game start", "level completed" and "game end" can easily be tracked. An event represents a code-path being triggered, but the various behaviors that want to be studied may be described by several events happening in a particular sequence or context [8]. Raw telemetry data can be transformed into various interpretable measures referred as gameplay metrics. Behaviors that are analyzed may be described by several events happening in a particular sequence or context, so it is useful to be able to tie events together through common identifiers [8].

The focus of this paper is on building an extended RFM (Recency, Frequency and Monetary) prediction model for churn classification in the field of mobile games by using telemetry data. The RFM model is extended by incorporating features related to user lifetime, usage intensity and rewards. The analysis addresses the main problem, establishing a churn prediction model, but also several common issues in churn classification: constructing meaningful features on the basis of raw telemetry data, handling sparse data, assessing feature importance in the logit model, building a prediction model with imbalanced datasets and assessing the model fit. This work relies on using logistic regression for classifying users as churners, because the use of this method in many applications has shown that it offers a good predictive performance, understandable models, and interpretable log odds ratio to assess the explanatory variables. Moreover, logit models have several desirable features: logit modelling is well-known, conceptually simple and frequently used in marketing, especially at the level of the individual consumer; the ease of interpretation of logit is an important advantage over other methods (e.g. neural networks); logit modeling has been shown to provide good and robust results in general comparison studies for churn prediction and it has been shown by several authors that logit modeling may even outperform more sophisticated methods [5]. Correspondingly, logistic regression is frequently used for customer churn prediction and proved to be an efficient and accurate classifier [17, 22, 24].

2. Related work

Churn has been analyzed in a range of industries, most widely in telecommunication sector, but also in the field of banking and insurance, retail business, internet service providers, service industries, newspaper publishing, P2P networks and online games [14, 15]. Among many techniques for customer churn prediction modeling that have been proposed over the years, decision trees, neural networks and logistic regression stand out as the most popular techniques [12, 14, 23]. Other frequently used approaches include support vector machines, Naive Bayes, random forests, regression forests, different regression analysis, clustering, ensemble methods and many more. This plethora of statistical/data mining/machine learning approaches make it difficult to choose an appropriate model for churn modeling. Considering that the choice of the method affects the churn modeling, several authors gave guidelines for companies that are starting up a predictive modeling function. For instance, Neslin et al. [22] proposed logit and tree approaches as a good starting point. However, in spite of the fact that churn prediction modeling has been extensively researched, no general consensus exists on the performance

of churn prediction modeling techniques, and there are contradictory results when comparing such models [31]. Since conflicts often arise when comparing different customer churn prediction models, the issue of which technique to use remains an open research issue [30].

First two studies investigating churn prediction in F2P games were conducted in 2014 [13, 26]. Hadiji and co-authors defined a range of features (e.g. playtime, session length, and session intervals) that are generic to games, and evaluated their usefulness in predicting player churn [13]. They compared a range of different classifiers, including neural networks, logistic regression, naive Bayes and decision trees. They found that in terms of weighted F1-score, decision trees performed best, followed by logistic regression. Moreover, they showed that usage frequency, usage intensity and user lifetime stand out as features with the highest importances for churn prediction. Runge and co-workers [26] focused on predicting churn for high value players of casual social games. They compared optimal prediction performance of 4 common classifiers and showed that logistic regression and neural networks outperform decision trees and support vector machines. Kim and co-workers provided new insights by focusing on new players [18]. They formally defined churn by using different observation churn prediction periods. For churn prediction modeling, they utilized logistic regression, gradient boosting, random forests and two deep learning algorithms, CNN and LSTM. Selection of the optimal algorithm varied over the games that were included in analysis. Still, since the performance advantage of the best performing algorithm was not very large compared to any other algorithm, they conclude that the prediction performance has a little dependency on the choice of the machine learning algorithm. They also showed that playtime related metrics matter much more than play-score related metrics for casual games. Casual game players tend to churn very quickly and for this reason, a challenging problem is to detect churn at an early stage. For instance, Milošević et.al [21] explored early one-day churn prediction and applied four machine learning algorithms for the binary classification problem of churn prediction: logistic regression, decision trees, random forest, gaussian naive Bayes and gradient boosting. Tree-based models mostly outperformed all other models in terms of AUC and F1 measures where the gradient boosting model performed best, while decision trees showed the worst results. Logistic regression has shown good performance, considering the simplicity and scalability of the model. Very short-term user behavior was evaluated by Drachen and co-workers [7] and they revealed that retaining players can be successfully determined with a short history of behavioral information and using heuristic prediction approaches. Moreover, they suggest that a large part of the value of advanced analytics in games can potentially be accessed by relying on static heuristic models because these models are robust, easily understandable and simple to deploy and scale. Several studies have moved fom the traditional binary classification to evaluating time-to the event, i.e. time to churn. For instance, Perianez and co-workers [25] modeled the churn behavior for high value players from the perspective of survival analysis. They treated the churn prediction as a problem of censored data,where the outcome is the continuous time – the time-to-exit the game. Viljanen and co-workers [32] analyzed player retention as a duration variable and used total playtimes of an in-development mobile game with the application of survival analysis. They emphasize that retention can be quantified as user engagement by employing several duration variables: playtimes, session lengths, subscription times, and game progression.

3. Methodology

3.1. Data collection

This study builds on the data that was collected in 2018, using randomly selected players of a F2P online casual game developed and published by Nanobit, Croatia. Dataset was constructed to simulate a business problem where on a specific date, a selection of active users would be used to evaluate their churn risk by analyzing their (past) behavior. Active users that are

at a high risk level could potentially be incentivized to stay in the game by involving them in different retention campaigns. Here, the target user population is the population of active users, regardless their lifetime. The dataset was formed by randomly selecting 10 000 users that had an event recorded on a certain date D_1. In order to eliminate users that have no observable past activity and engagement, filtering procedures were performed. If users that just registered, but do not have additional sessions, were kept in the sample, this would skew many variables, such as user activity or spending and would generate samples with no history observations to be analyzed. For this reason, users that have experienced the game with at least one session, one level completed and at least 5 minutes total session duration were kept in the dataset. Focusing not solely on events recorded, but on usage duration and levels completed is considered as needful due to the fact that event recording does not necessary imply experiencing the game. For instance, a user can start a game, and leave it opened while actually doing something else. This filtering procedure eliminated around 16 percent of users from the dataset, where 11 percent were users with no observable past. This left 8415 users and about three million events in the sample. During data preparation and, especially, during data cleansing, inaccurate records were deleted or modified. The main data cleansing procedure refers to calculation of session duration and number of sessions. Session duration was calculated by evaluating two events: "game start" and "game end", and parameters that are recording time of these events. Session duration is defined as a time difference between two consecutive "game start" and "game end" events, where an upper bound on session duration of 10800 seconds was set. Thus, session durations over 10800 seconds were limited on 10800 seconds. The second modification was set on calculating the number of sessions. In order to remove irrelevant sessions, when calculating the number of sessions, only sessions with duration 60 seconds or more were taken into account. During the observation period of 1 week, tracked events were aggregated on a daily basis for each user.

3.2. Data transformation

The initial step in developing a customer churn management model is identifying "the best data". The quality of the data used in modeling will determine the power and the accuracy of the model. Since different combinations of data have different analytical power, it is necessary to determine most suitable data for the analysis being performed. RFM features (Recency, Frequency and Monetary) proved to be a good source of customer behavior in general [3, 12], but also in the F2P online gaming market [6, 27]. Recency features are related to the time of the last usage of the service; frequency features are related to how often the service is used, while monetary features are related to the total money that the customer has spent on services over a certain period. In the mobile gaming field, information on user behavior is found in player's action log, daily logins, purchases, playtime, and level-ups [2] and there is a simple way to transform these events into features. For instance, Castro and Tsuzuki propose an approach where recency would represent a time span since the player's last game session; frequency would correspond to the number of game sessions recorded during a predetermined period of time [6].Moreover, if no monetary value is directly involved, they propose replacing the third metric by some measure of "intensity" of a past activity such as the sum of all session times recorded from the player's past activity. In this research, the RFM framework is extended with additional features, related to user lifetime, rewards and usage intensity. In this case, usage intensity observes user activity and engagement relative to user lifetime. As mentioned before, one of the benefits of the RFMs approach is its simplicity. However, this simplicity is also one of the major drawbacks, because it can yield exactly the same feature values for remarkably different playing behaviors [6]. By aggregating (summing) feature values through the observation period the time aspect is completely neglected.

In this paper, features are derived from six gameplay events: "game start", "game end",

"level-up", "currency spent", "purchase completed" and "commercial video watched". For each event, different parameters, such as timing and different currency amounts, were being tracked. Based on these events, a set of features useful for churn prediction was established and presented in Table 1.

Feature category	Feature [feature label] # selected for churn model	Transformation	Recording frequency
Frequency	#Level completed [Lvl_Nr] #Session number [Ses_Nr] Session length [Ses_Len]	EMA	Daily
	Total lifetime session length [$LifeSes\_Len$] Total lifetime session number [$LifeSes\_Nr$] #Total lifetime levels completed [$LifeLvl\_Nr$]	-	End of observation period
Lifetime	#Lifetime in days [$Life$]	-	End of observation period
Monetary (currency spent and purchases)	#Currency 1 spent [$C1\_Spnt$] Currency 2 spent [$C2\_Spnt$]	EMA	Daily
	#LTV (lifetime value) [LTV]	Discretized (payer/ non-payer)	End of observation period
Rewards (monetary & non-monetary)	#Commercial watched –currency reward [Rew_sec]	Discretized (watcher/ non-watcher)	Daily
	#Commercial watched – rewards in seconds [Rew_cur]	Sum	Daily
Recency	#Days from last session [Rec_Ses] #Days from last level completed [Rec_Lvl]	Discretized (having/not having session/level completed yesterday)	End of observation period
	#Number of days from last currency 1 spent [Rec_C1]	-	End of observation period
	#Number of days from last currency 2 spent [Rec_C2]	Discretized (having/not having currency2 spent yesterday)	End of observation period

Table 1: *Constructed features*

Features are constructed on the basis of similar research in accordance with data scientists from the developer company Nanobit. In order to aggregate a series of daily data recorded during one week into a single indicator while simultaneously incorporating the time effect, Exponential Moving Average (EMA) was applied. Due to the fact that some users had lifetime length less than one week, exponential moving average was applied conditional on user lifetime as defined with:

$$EMA_t^x(i) = (x_t(i) - EMA_{t-1}^x(i))\frac{2}{k+1} + EMA_{t-1}^x(i), \quad k = \min(6, lifetime(i)), \quad (1)$$

where $x_t(i)$ is the value of feature x of user i on day t, $EMA_t^x(i)$ is the value of exponential moving average value calculated for feature x of user i for day t, where 6-day exponential moving average was applied if user lifetime is larger than 6, while otherwise the time period of EMA is limited to user lifetime length. Exponential moving average is weighted toward the

most recent data and older data become exponentially less important in the average calculation. Based on features presented in Table 1, three additional features were constructed: mean daily session length [$ADay\_Ses\_len$], mean daily session number [$ADay\_Ses\_Nr$], mean daily levels completed [$ADay\_Lvl\_Nr$]. These features relate three frequency features (total lifetime session length, total lifetime session number, total lifetime levels completed) to user lifetime in days, and as such can express usage intensity. Some of the features were discretized due to high skewness or high sparsity (Table 1). For instance, lifetime value LTV was transformed into a binary feature that distinguishes payers from non-payers. Selection of features that will be used for churn prediction was based on feature reliability, which was discussed with the developer, and examined with the analysis of the raw dataset. Another important parameter for the selection of features was their theoretical and empirical relation with other features. Since the construction of all features was based on a raw telemetry dataset that consists of recorded events, theoretical relation stems from the pure feature design. Empirical relation, on the other hand, was determined through correlation analysis. For example, EMA transformed session length [Ses_Len] is highly correlated with EMA transformed level completed [Lvl_Nr] and due to the fact that capturing and explaining levels completed has a higher reliability than capturing number of sessions, only Lvl_Nr was used for churn prediction modelling.

3.3. Churn definition

In non-contractual business settings, such as freemium mobile applications and online games, customers do not churn through the termination of a contract. Leaving the game or an app is very easy for players without notifying the firm about their decision. In the field of mobile games, churn definition usually relies on player absence in an a priori selected period of time. A player is usually classified as a churner if she/he is not using a game within a given period of time. Decision of whether a customer has churned or not is quite subjective, and the churn definition is usually based on heuristic rules set by the industry officials [16]. Thus, churn definitions are highly influenced by the subjective perception of the modeller who has to decide (inter alia) on the best suited behavioral measure to evaluate and prediction window size. Duration of the prediction period may vary between applications, depending on different business goals, where 14 days and 30 days are some typical settings used in industry [19]. This paper focuses on the engagement-based approach and identifies churners on the basis of their lack of engagement in a game through completing a level in a game, not solely by their absence. Thus, players are identified as churners if they have no level completed in the period of 14 consecutive days.

3.4. Building churn prediction model and assessing feature importance

First, features were analyzed based on the notion of explanatory power. The analysis was based on robust statistical methods that can overcome several issues: high asymmetry level, outlier presence and heteroscedasticity between churn/non-churn groups. Second, the difference between churners and non-churners was estimated by calculating robust effect sizes proposed by Wilcox and Tian [33]. As last, calculated effect sizes were used to rank the features.

Before building the churn prediction model, the dataset was randomly split into train and test datasets whereas 75 percent of the data were used for training the model. Churn problems usually introduce imbalanced datasets. This can be handled by different techniques where most common techniques rely on sampling and follow the basic principle of minimizing or even eliminating rarity by altering the distribution of training examples [4]. In this work, the problem of imbalanced dataset was handled by performing random undersampling approach, which eliminates majority-class examples (non-churners) and results in a completely balanced dataset.

Strategy for logistic model building relied on identifying the most parsimonious model in

order to obtain a numerically stable and more easily generalized model. Stepwise logistic regression with forward and backward elimination was applied, choosing the best model according to the Akaike information criterion (AIC). The model fit was assessed by performing Hosmer-Lemeshow goodness of fit and by the measures derived from classification tables: accuracy, sensitivity (recall) and by calculating AUC. A special emphasis was put on the sensitivity measure and AUC, due to the fact that classification accuracy can be misleading in the case of imbalanced datasets.

The importance of each feature in the process of modeling the churn prediction was examined by applying dominance analysis, which is an intuitive and a straightforward procedure for determining predictor importance. This type of analysis requires only a measure of model fit to determine an additional contribution of any given predictor to any specific subset model, and follows the principle that "one predictor is more important than another only if it would be chosen over its counterpart in all subset models" [1]. More specifically, dominance analysis used in this work relies on conditional and general dominance, which are weaker dominance types and operate on an "on average" basis. General dominance is calculated as an average of each feature conditional dominance measures, and features were then ranked according to their general dominance. The criterion applied was McFadden's R^2.

4. Results

4.1. Churn vs. non-churn group

The engagement churn definition given in 3.3 classified 15 percent of users as churners. For all examined features, churners and non-churners exhibit significant differences with mostly high or medium effect sizes as presented in Table 2 where 0.15, 0.35, and 0.50 correspond to small, medium and large effect sizes, respectively [33]. Long term frequency features have the largest effect sizes, followed by short term frequency, intensity and monetary features. Compared to non-churners, churners exhibit lower mean values of all quantitative features except for number of days from last currency 1 spent.

LifeLvl_Nr	LifeSes_Nr	LifeSes_Len	Ses_Nr	ADay_Ses_Nr	C2_Spnt	Lvl_Nr	Ses_Len	Rec_C1	ADay_Ses_Len	Rec_C2	ADay_Lvl_Nr	Rec_Lvl	Rec_Ses	Rew_sec	LTV	C1_Spnt	Life	Rew_cur
0.45	0.45	0.44	0.39	0.38	0.38	0.37	0.36	0.35	0.33	0.33	0.33	0.32	0.31	0.28	0.12	0.12	0.07	0.04

Table 2: *Feature effect sizes*

4.2. The logit model

In order to completely balance churn and non-churn classes, random undersampling was performed on the train data. This way, the training dataset was formed by taking a random sub-sample of non-churners from the initial training dataset. Importance of the constructed features was evaluated by applying the stepwise logistic regression with forward and backward elimination to several undersampled train datasets. Only significant features were then selected for examining possible intercorrelations. This was achieved by performing stepwise logistic regression with forward and backward elimination and all possible intercorrelations included. This process was repeated after excluding outliers determined on the basis of Cook's distance. Model results are presented in Table 3.

Feature	Coefficients	S.E.
Intercept	1.1670 ***	0.184
C1_Spnt	0.1829 **	0.066
Lvl_Nr	−0.6185 ***	0.183
LifeLvl_Nr	−0.0055 **	0.002
ADay_Ses_Nr	−0.1165 **	0.045
ADay_Lvl_Nr	0.2260 *	0.093
Rec_C1	0.0258 ***	0.008
LTV	0.0189	0.249
Rec_C2	−0.8436 ***	0.206
Lvl_Nr×LifeLvl_Nr	−0.0025 ***	0.001
Lvl_Nr×Rec_C1	−0.0190 **	0.007
ADay_Ses_Nr×ADay_Lvl_Nr	0.0164 **	0.005
C1_Spnt×Rec_C2	−0.1766 **	0.066
Lvl_Nr×Rec_C2	0.6546 ***	0.182
ADay_Lvl_Nr×Rec	−0.2612 **	0.098
ADay_Ses_Nr×LTV	−0.1572 **	0.060
C1_Spnt×LifeLvl_Nr	0.0003	0.001

Table 3: *Logit model results (***, **, * indicate significance at the 1%, 5% and 10% levels)*

Hosmer–Lemeshow goodness of fit suggests that there is no (strong) evidence on the model lack of fit ($\chi^2 = 12.7$, $df = 8$, $p-value = 0.12$). According to classification matrix the accuracy of the model is 71%, while the sensitivity reaches 73% (Table 4). The ROC curve is presented on Figure 1 with AUC = 79%. Model performance was evaluated on the test data with the accuracy of 67% and sensitivity of 73%.
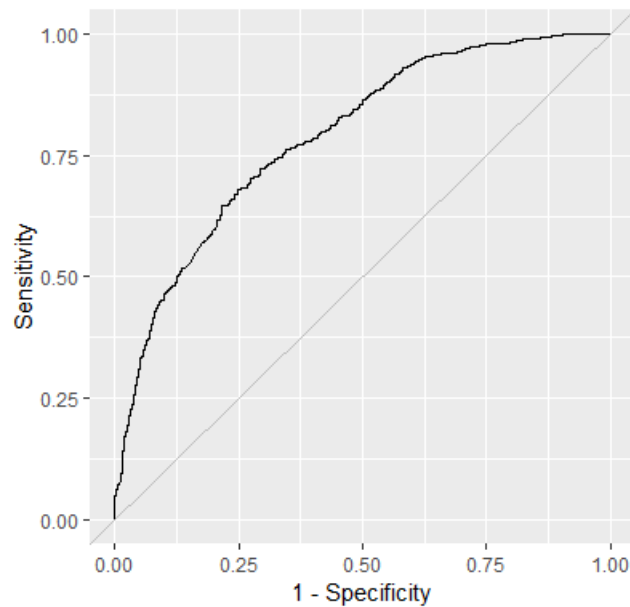
Figure 1: *ROC curve*

Predicted / Actual	Non-churn	Churn
Non-churn	647	256
Churn	288	688

Table 4: *Classification matrix*

Logistic model was benchmarked against random forest model based on unbiased trees that enables learning unbiased forests [28, 29]. The model was built after hyperparameter tuning on a validation dataset by a grid search and maximizing classification sensitivity. Random forest model achieved lower values of sensitivity (72%) and AUC (72%), but ahigher accuracy level (72%). Stepwise logistic regression was also performed on the overasampled train dataset. Significant features for the stepwise model without interactions were the same as in the undersampled case, but the model with intercorrelations included had higher complexity and a lower performance.

4.3. Feature importance

Dominance analysis was applied in order to determine importance of features in the churn prediction model. The importance is firstly asserted separately for different model sizes whereas the model size indicates the number of features included in the model. This was performed firstly by calculating additional contribution of each feature in terms of McFadden's R^2 for different models of the same size, and then averaging calculated values for each feature. The results are presented by a heatmap on Figure 2.
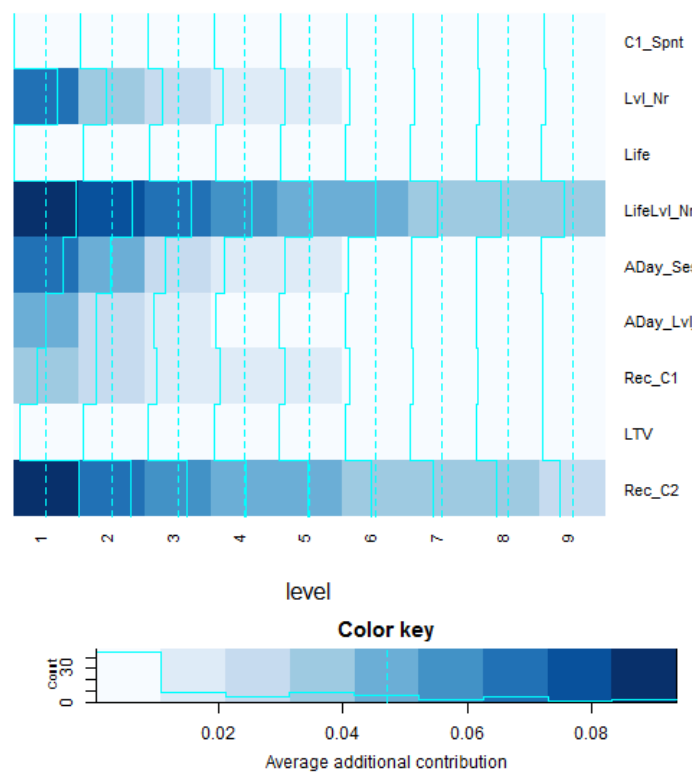

Figure 2: *Average additional feature contribution by level*

Furthermore, general feature dominance is determined by calculating the average of conditional measures for each feature. This is presented in Table 5 where features are ranked according to their importance. Feature importance ranks are consistent with the robust univariate effect size results examined prior to building the prediction model. $LifeLvl\_Nr$ and Rec_C2 stand out as the most important features generally dominating other features included in the churn prediction model.

LifeLvl_Nr	0.055
Rec_C2	0.05
ADay_Ses_Nr	0.021
Lvl_Nr	0.021
Rec_C1	0.014
ADay_Lvl_Nr	0.012
C1_Spnt	0.005
LTV	0.004
Life	0.004

Table 5: *General feature dominance*

5. Conclusion

As mobile game markets became increasingly saturated, companies have recognized that increasing the revenue is more likely to come from their existing customer base than from a new acquisition. One way of preventing the customer churn is to predict it, so retention management in the gaming industry has shifted its focus the customer churn prediction. Accurate prediction of customer churn is an ongoing challenge for both gaming industry and academia.

This paper presents a procedure for building a customer churn prediction logit model within the extended RFM framework for the mobile gaming market. The first step of the procedure focuses on transforming the raw telemetry data into meaningful game independent features that can be used for building the churn prediction model within the extended RFM framework. Feature univariate exploratory power is evaluated by applying a robust heteroscedastic approach. This approach is capable of handling skewed, heavy-tailed features generated from the raw telemetry data. Based on selected features, a logistic churn prediction model is built where the feature importance is assessed by dominance analysis. The established logit churn prediction model proved to be an efficient and accurate classifier, and was benchmarked by a random forests model.

Besides solely churn prediction, understanding why users churn is another challenging task in the development of retention campaigns. Consequently, the most important benefit of the established extended RFM logit model is related to interpretability of constructed features and classification results. Our analysis showed that users classified as churners tend to have lower long term and short term engagement and activity, lower engagement intensity and usage intensity, they spend less, and are less interested in performing additional activities in order to receive rewards. Also, churners have lower median lifetime than non-churners. Differences in recency depend on the feature examined where the proportion of non-churners that recently have spent currencies or have completed at least one level, tend to be higher than the proportion of non-churners. On contrary, the proportion of recently active churners is higher than the proportion of recently active non-churners. The magnitude of these univariate differences is examined by analyzing robust effect sizes. The results showed that the exploratory power is

highest for the long-term frequency features, followed by short-term frequency features, intensity features and monetary features. Similar results are obtained when examining the feature importance for churn prediction models that involve feature combinations. Within the multivariate logit churn prediction model, long term frequency features and recency features stand out as the most important features for churn prediction, followed by intensity features and short term frequency features.

Within the vast amount of information and possible reasons that lead to mobile game abandonment, it is crucial to recognize the most important features, especially since some features are more important than the others in terms of their contribution to churn prediction. Studies on churn prediction in mobile games have involved various game-dependent and game-independent features for churn prediction, applied different churn definitions, and utilized different prediction algorithms for different subpopulations of users. Features were rarely categorized into meaningful groups by their characteristics or on the basis of some framework, and reporting feature importance is not a norm. Thus, results in terms of the model predictive performance or feature importance may not be directly comparable. Still, many features can be categorized into the extended RFM framework for achieving feature importance comparability. Within studies that have reported feature importance, frequency features were regularly reported as one of the most important features [7, 13, 18, 25]. Also, recency, lifetime, monetary and intensity features were distinguished as features with the highest importance [7, 13, 18, 25]. Moreover, intersession measures related to playtime consistency, such as intersession time, were also reported as highly important [7, 13]. This opens new challenges in extending the proposed feature framework.

Two main limitations can be singled out within this study: (1) the results presented in this paper are based on data coming from only a single game; (2) the results presented are based on a single, arbitrarily chosen churn definition. The second issue is a common situation for churn prediction analysis in non-contractual settings where churn definitions are highly influenced by the subjective perception of the modeller. The modeller has to decide on which behavioral measure and prediction window size to use. Although the proposed extended RFM logit model proved to be a very good discriminator between churners and non-churners, in order investigate the full potential of RFM and its extensions, different methods should be tested and different observation period length should be analyzed.

Acknowledgements

References

[1] Azen, R. and Traxel, N. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics*, 34(3), 319–347. doi: 10.3102/1076998609332754

[2] Bertens, P., Guitart, A. and Perianez, A. (2017). Games and big data: A scalable multi–dimensional churn prediction model. *IEEE Conference on Computational Intelligence and Games*, CIG 2017, 33–36. doi: 10.1109/cig.2017.8080412

[3] Buckinx, W. and Van Den Poel, D. (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non–contractual FMCG retail setting. *European Journal of Operational Research*, 164(1), 252–268. doi: 10.1016/j.ejor.2003.12.010

[4] Burez, J. and Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), 4626–4636. doi: 10.1016/j.eswa.2008.05.027

[5] Burez, J. and Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2), 277–288. doi: 10.1016/j.eswa.2005.11.037

[6] Castro, E. G. and Tsuzuki, M. S. G. (2015). Churn prediction in online games using players' login records: A frequency analysis approach. *IEEE Transactions on Computational Intelligence and AI in Games*, 7(3), 255–265. doi: 10.1109/tciaig.2015.2401979

[7] Drachen, A., Lundquist, E. T., Kung, Y., Rao, P., Sifa, R. , Runge, J. and Klabjan, D. (2016). Rapid prediction of player retention in free-to-play mobile games. *Proceedings of AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, AIIDE 2016, 23–29. https://www.aaai.org/ocs/index.php/AIIDE/AIIDE16/paper/download/13995/13590

[8] Drachen, A., and Connor, S. (2018). *Game analytics for games user research*. In Drachen, A., Mirza–Babaei, P. and Nacke, L. E. (Eds.) Games User Research. Oxford University Press. doi: 10.1093/oso/9780198794844.003.0019

[9] Drachen, A., Mirza–Babaei, P. and Nacke, L. E. (2018). *Frontlines in games user research*. In Drachen, A., Mirza–Babaei, P. and Nacke, L. E. (Eds.) Games User Research. Oxford University Press. doi: 10.1093/oso/9780198794844.003.0031

[10] Drachen, A., Seif El–Nasr, M. and Canossa, A. (2013). *Game analytics–the basics*. In Seif El–Nasr, M., Drachen, A. and Canossa, A. (Eds.) Game Analytics (pp. 13–40). London: Springer. doi: 10.1007/978-1-4471-4769-5_2

[11] Fields, T. V. (2013). *Game industry metrics terminology and analytics case study*. In Seif El–Nasr, M., Drachen, A. and Canossa, A. (Eds.) Game Analytics (pp. 53–71). London: Springer. doi: 10.1007/978-1-4471-4769-5_4

[12] Hadden, J., Tiwari, A., Roy, R. and Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers and Operations Research*, 34(10), 2902–2917. doi: 10.1016/j.cor.2005.11.007

[13] Hadiji, F., Sifa, R., Drachen, A., Thurau, C., Kersting, K. and Bauckhage, C. (2014). Predicting player churn in the wild. *IEEE Conference on Computatonal Intelligence and Games*, CIG 2014, 1–8. doi: 10.1109/cig.2014.6932876

[14] Hashmi, N., Butt, N. A. and Iqbal, M. (2013). Customer churn prediction in telecommunication: A decade review and classification. *International Journal of Computer Science Issues*, 10(5–2), 271–282. https://ijcsi.org/papers/IJCSI-10-5-2-271-282.pdf

[15] Karnstedt, M., Hennessy, T., Chan, J., Basuchowdhuri, P., Hayes, C., and Strufe, T. (2010). *Churn in social networks*. In Furht, B. (Eds.) Handbook of Social Network Technologies and Applications (pp. 185–220). Boston: Springer. doi: 10.1007/978-1-4419-7142-5_9

[16] Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., Bozkaya, B. and Pentland, A. S. (2018). Behavioral attributes and financial churn prediction. *EPJ Data Science*, 7(1), 41. doi: 10.1140/epjds/s13688-018-0165-5

[17] Kim, H. S. and Yoon, C. H. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28(9–10), 751–765. doi: 10.1016/j.telpol.2004.05.013

[18] Kim, S., Choi, D., Lee, E. and Rhee, W. (2017). Churn prediction of mobile and online casual games using play log data. *PLOS ONE*, 12(7), e0180735. doi: 10.1371/journal.pone.0180735

[19] Liu, X., Xie, M., Wen, X., Chen, R., Ge, Y., Duffield, N. and Wang, N. (2018). A semi–supervised and inductive embedding model for churn prediction of large–scale mobile games. *IEEE International Conference on Data Mining*, ICDM 2018, 277–286. doi: 10.1109/icdm.2018.00043

[20] McAllister, G. and Long, S. (2018). *A framework for player research*. In Drachen, A., Mirza–Babaei, P. and Nacke, L. E. (Eds.) Games User Research. Oxford University Press. doi: 10.1093/oso/9780198794844.003.0008

[21] Milošević, M., Živić, N. and Andjelković, I. (2017). Early churn prediction with personalized targeting in mobile social games. *Expert Systems with Applications*, 83, 326–332. doi: 10.1016/j.eswa.2017.04.056

[22] Neslin, S. A., Gupta, S., Kamakura, W., Junxiang, L. U. and Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, 43(2), 204–211. doi: 10.1509/jmkr.43.2.204

[23] Ngai, E. W. T., Xiu, L. and Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592–2602. doi: 10.1016/j.eswa.2008.02.021

[24] Owczarczuk, M. (2010). Churn models for prepaid customers in the cellular telecommunica-

tion industry using large data marts. *Expert Systems with Applications*, 37(6), 4710–4712. doi: 10.1016/j.eswa.2009.11.083

[25] Perianez, A., Saas, A., Guitart, A. and Magne, C. (2016). Churn prediction in mobile social games: Towards a complete assessment using survival ensembles. *IEEE International Conference on Data Science and Advanced Analytics*, DSAA 2016, 564–573. doi: 10.1109/dsaa.2016.84

[26] Runge, J., Gao, P., Garcin, F. and Faltings, B. (2014). Churn prediction for high–value players in casual social games. *IEEE Conference on Computatonal Intelligence and Games*, CIG 2014, 1–8. doi: 10.1109/cig.2014.6932875 https://doi.org/10.1109/CIG.2014.6932875

[27] Sifa, R., Drachen, A. and Bauckhage, C. (2018). *Profiling in games: Understanding behavior from telemetry*. In Lakkaraju, K., Sukthankar, G. and Wigand, R. T. (Eds.) Social Interactions in Virtual Worlds (pp. 337–374). Cambridge University Press. doi: /10.1017/9781316422823.014

[28] Strobl, C., Boulesteix, A. L., Zeileis, A. and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. doi: 10.1186/1471-2105-8-25

[29] Strobl, C., Boulesteix, A.–L., Kneib, T., Augustin, T. and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307. doi: 10.1186/1471-2105-9-307

[30] Verbeke, W., Dejaeger, K., Martens, D., Hur, J. and Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1), 211–229. doi: 10.1016/j.ejor.2011.09.031

[31] Verbeke, W., Martens, D., Mues, C. and Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3), 2354–2364. doi: 10.1016/j.eswa.2010.08.023

[32] Viljanen, M., Airola, A., Heikkonen, J. and Pahikkala, T. (2018). Playtime measurement with survival analysis. *IEEE Transactions on Games*, 10(2), 128–138. doi: 10.1109/tciaig.2017.2727642

[33] Wilcox, R. R. and Tian, T. S. (2011). Measuring effect size: A robust heteroscedastic approach for two or more groups. *Journal of Applied Statistics*, 38(7), 1359–1368. doi: 10.1080/02664763.2010.498507