Predicting Churn Rate Of The Massively Multiplayer Online Role-Playing Game (Mmorpg) Users By Analyzing Playing Behavior

Han-Soon Sin, Woojin Paik

Abstract: This study examined various methods to predict when the Multiplayer Online Role-Playing Game (MMORPG) players might quit the game. A relatively large data set consisting of 100,000 cases and 37 predictors was used. The data set was visualized with boxplots to conduct the exploratory analysis. Binary logistic regression and multi-layer perceptron were used to develop a prediction model. The boxplot analysis revealed that players who engage less in social interactions were likely to quit the game. This finding was partially confirmed by binary logistic regression analysis. Only the permanent form of chat type significantly affected user churn. The multi-layer perceptron model slightly outperformed the binary logistic regression in terms of the prediction accuracy at 85%.

Index Terms: binary logistic regression, boxplot visualization, multi-layer perceptron, online game, player behavior, prediction, user churn

1. INTRODUCTION

ONLINE games are a vital part of the entire gaming industry today with a continuous growth curve. In 2018, the global game industry revenue was estimated to be around USD 135 billion, and it is expected to grow to USD 180.1 billion by 2021 [1]. Massively Multiplayer Online Role-Playing Game (MMORPG) is one of the online game genres enjoyed by many users around the world. MMORPG is a role-playing game played by tens to hundreds, and sometimes thousands of players in the same virtual space. Some of the game users lose interest over time and quit playing. The resulting user churn directly affects the revenue of the game company and indirectly influence potential and existing users through word-of-mouth stories [2]. As the game industry grows in size, it is important for them to be able to anticipate and prevent user churn for the growth of both the company and profit [3]. A change in usage time and in-game behavior is noticed just before an online game user quits the game [4]. Thus, it will be possible to predict when the user quits the game by analyzing the recorded activities of the users. The findings will enable the game providers to entice potential quitters to continue to play the game. This study used game data about Blade & Soul, which is a fantasy MMORPG developed by NCSOFT, a South Korean video game developer. Blade & Soul features martial arts-based combats. Players explore around the world by completing various Non-Player Characters (NPCs) assigned quests [5]. There are 36 types of data recorded about every aspect of user actions in the game log — data about the quitters and the rest about the continuing players - each data type equated to a variable in the ensuing analysis. . The data was analyzed using visual inspection-based on the box plot graphs, churn prediction using logistic regression, and finally, churn prediction using deep learning algorithms. The primary goal of this study was finding who the quitters are based on the combination of 36 variables.

- *Han-Soon Sin completed masters degree program in computer engineering in Konkuk University Glocal Campus, South Korea, E-mail: qzwxec1994@naver.com*
- *Woojin Paik is a professor at Dept. of Software Technology at Konkuk University Glocal Campus, South Korea, E-mail: wjpaik@kku.ac.kr, Corresponding author*

The secondary goal was to find ways to prevent users from quitting. Various studies have been carried out with attempts to resolve user churn problems in online games. These studies can be analyzed in a variety of ways, depending on the data collection method. Automatically generated user log of every action taken by the players was used to extract the behavioral variables, which might affect the player's decision to either quit or continue playing. The variables were used to develop a churn prediction model. Bobora et al. [6] predicted churning of Sony Online Entertainment's EverQuest II game by using a log analysis-based ensemble technique. According to their model, feature variables related to achievement motivation had high predictive power. A study was carried out to develop a churn model based on the top-level players' behavioristic characteristics. Park and Cha [3] discovered that players with an elevated level of social interaction continue to play the game even when they achieved the highest possible level. However, social interaction did not always have a positive impact on continuing the game. Ducheneaut et al. [7] revealed that certain types of social interactions that made the cohesion within a guild weaker and affected the churn rate. "A guild is a group of players that decide to play together for a period exceeding the length of one playing session. [8]." Milosevic et al. [9] showed that sending push alarms to mobile social game players, who showed signs of quitting, reduced churn up to 28%.Some studies have been carried out predicting user churn using various sets of variables about players' actions within the online game. In this investigation, the variables were divided into four groups before the analysis: 1) raid related where multiple users perform one quest, 2) quest related, 3) chat related, and 4) combat-related variables. Raids and combats were expected to be critical game components of MMORPG, the genre of Blade and Soul. Therefore, a hypothesis was developed, stating that the players who enjoy the raids and combats will be less likely to quit playing the game.

2 METHODOLOGY

The online games record their players' actions as logs. Eight weeks' worth of log data of Blade & Soul received from NCsoft. During the eight weeks, 100,000 players and 440,324 activities were recorded. Any piece of data that might reveal the identity of the players, such as the user ids, party ids, and item ids, were hidden by NCsoft before releasing the data.

TABLE 1

PLAYER STATUS WITH FOUR POSSIBLE VALUES

Variable	Definition and possible value
acc_id	account id
week	the player quit between the eighth week and the ninth week
month	the player quit between the ninth week and the 12th week
2month	the player quitting the game sometime between the 12th week and the 16th week
retained	the player had not quit for more than eight weeks after the data collection period

The data set included six files. One had player status with four possible values concerning when she quit playing the game (Table 1). Another file aggregated each user's in-game activities weekly within the eight-week long data collection period. There were 37 activities, which in turn are 37 variables and two non-activity related variables (i.e., week_id and acc_id) in Table 2. Another file showed the party membership (Table 3). The party in MMORPG is defined as "players team up and go on raids against another team [10]." Another file

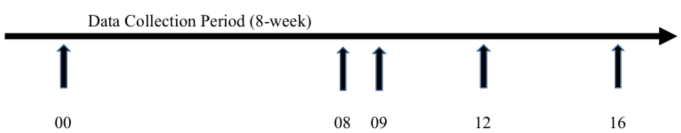

Fig. 1. *Timeline affecting the Player Status starting from the Data Collection 8-week Period*

was about the guild membership (Table 4). "Guilds, also known as clans, kinships, or crews, are usually groups of people who are working toward a common goal in a game [11]. Another file had the transaction history among the players (Table 5). Finally, the sixth file showed each player's payment for purchasing goods or services information by week (Table 6). All numerical data were standardized using z-score normalization [12] as the first step of processing after receiving the data.There were four possible player status. Fig. 1 shows the timeline affecting the status and the beginning of the eight-week-long data collection period as the zeroth week. The 'retained' status means that the player had not quit for more than eight weeks after the data collection period or the player not quitting the game until the end of the 16th week in the figure. The one-week churn status means that the player quit between the eighth week and the ninth week. The four-week churn status means that the player quit between the ninth week and the 12th week. Finally, the eight-week churn status means that the player quit between the 12th week and the 16th week. 25,000 player data were recorded for each status. For each variable in Table 2, descriptive statistics were computed to get the overall sense of the data regardless of the player status.Some players' data did not include all activities during the eight-week data collection period. For example, there might be only records from the fifth week of the data collection period for a player, who belonged to the four-week churn status. Thus, it was decided that the data from the eighth week of the data collection period for the analysis be used because all players had the eighth-week data regardless of their churn status.

TABLE 2

EACH USER'S AGGREGATED IN-GAME ACTIVITIES BY WEEK

Variable	Definition and possible value
week_id	week when the activities occurred (1-8)
acc_id	account id
cnt_dt	number of days that the player logged in (1-7)
play.time	playing time (in seconds)
npc.exp	hunting Non Player Character (NPC) general experience level
npc.hongmun	hunting Non Player Character (NPC) Hongmum (practice arena) experience level
quest.exp	quest completion general experience level
quest.hongmun	quest completion Hongmum (practice arena) experience level
item.hongmun	item gathering Hongmum (practice arena) experience level
game.combat.time	combat time (in seconds)
get.money	amount of money acquired
duel.cnt	duels participated
duel.win	duels won
partybattle.cnt	battles participated as a member of a party
partybattle.win	battles won as a member of a party
cnt.enter.inzone.solo	entering inzone as solo
cnt.enter.inzone.light	entering inzone as light
cnt.enter.inzone.skilled	entering inzone as skilled
cnt.enter.inzone.normal	entering inzone as normal
cnt.enter.raid	participating a raid
cnt.enter.raid.light	participating a light raid
cnt.enter.bam	entering bam
cnt.clear.inzone.solo	clearing inzone as solo
cnt.clear.inzone.light	clearing inzone as light
cnt.clear.inzone.skilled	clearing inzone as skilled
cnt.clear.inzone.normal	clearing inzone as normal
cnt.clear.raid	clearing a raid
cnt.clear.raid.light	clearing a light raid
cnt.clear.bam	clearing bam
normal.chat	conducting normal chat
whisper.chat	conducting whisper chat
district.chat	conducting district chat
party.chat	conducting party chat
guild.chat	conducting guild chat
factions.chat	conducting faction chat
cnt.use.buffitem	using buff item
gathering.cnt	gathering items
making.cnt	making items

TABLE 3

PARTY INFORMATION AND MEMBERSHIP

Variable	Definition and possible value
party_start_week	week when the party was formed (1-8)
party_start_day	day when the party was formed (1-7)
party_start_time	time when the party was formed (00:00:00 ~ 23:59:59)
party_end_week	week when the party was abandoned (1-8)
party_end_day	day when the party was abandoned (1-7)
party_end_time	time when the party was abandoned (00:00:00 ~ 23:59:59)
party_members_acc_id	party members' account ids

TABLE 4

GUILD MEMBERSHIP

Variable	Definition and possible value
guild_id	Guild's unique id
guild_members_acc_id	guild members' account ids

TABLE 5

TRANSACTION HISTORY AMONG THE PLAYERS

Variable	Definition and possible value
trade_week	week when the transaction occurred (1-8)
trade_day	day when the transaction occurred (1-7)
trade_time	time when the transaction occurred (00:00:00 ~ 23:59:59)
source_acc_id	account id of the player who provides an item
target_acc_id	account id of the player who receives an item
item_type	money (gold) grocery weapon costume gem accessory
item_amount	number of items transferred

TABLE 6

PAYMENT FOR PURCHASING GOODS OR SERVICES

Variable	Definition and possible value
payment_week	week when the payment occurred
acc_id	account id of the payer
payment_amount	total amount of payment for the week

3 RESULTS

The primary goal of this study was to develop a model for predicting when a person might quit playing the online game. Initially, boxplots of each variable were drawn to get the feel of the data. Then, a logistic regression model was built to predict the user churn. Finally, a simple deep learning model was trained to predict the user churn then compared the prediction accuracy with the results from the logistic regression model.

3.1 Boxplot Visualization

For This was used to determine whether there are any differences in data distribution by player status. As shown in Table 1, there are four types of churn status based on when the players quit the game after the data collection period: 1) week, 2) month, 3) 2month, and 4) retained. Each activity-related variable in Table 2 was visualized as a boxplot using Tableau Public, which is a web service for creating an interactive visualization [13].

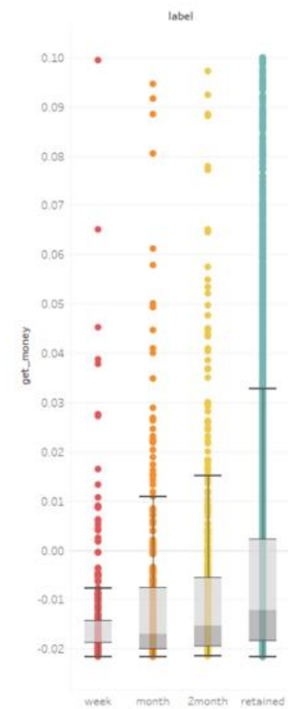

***Fig. 2**. Boxplot of the amount of money acquired (get_money) variable against four churn status*

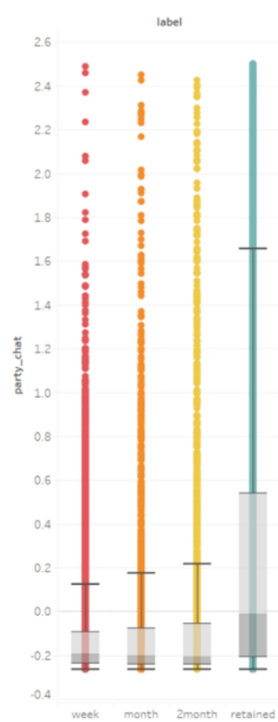

***Fig. 3.** Boxplot of the number of times conducting party chat (party_chat) variable against four churn status*

Fig. 2 shows the amount of money acquired (get_money) variable plotted against four churn status. It shows the players who stayed had the most money earned, and the ones who left the game after a week had the least with respect to the median value. The players who invested more played longer. Fig. 3 shows the number of times conducting party chat (party_chat) variable plotted against four churn status. It shows the players who stayed longer had the most frequent party chat, and the ones quit the game after a week had the least with respect to the median value. Players who formed the parties more and talked more played longer. Fig. 4 reveals a similar pattern for the number of times conducting guild chat (guild_chat). It shows that the players who belong to a guild have more attachment to the game and play longer. The boxplot analysis identified the variables that influenced strongly in differentiating the retained from the ones who quit as: playing time, frequency of chat, and the amount of money spent. In terms of chatting, the ones who quit within one week after the data collection period had the most frequent general chatting - i.e., talking with anyone nearby while playing the game. However, the retained players talked most frequently amongst the parties and guilds. In summary, the players who actively engaged in battles or duels but had a few social interactions were more likely to quit the game after a while. The ones who enjoy social interactions and being a part of the groups kept on playing the game. Thus, it might be possible for the game providers to reduce churning by offering incentives to the players to engage more with other players. The boxplot analysis provided a qualitative view of the phenomenon of game player churn. The logistic regression and deep-learning-based analysis is a more precise way to predict the game player churn.

3.2 Logistic Regression

"Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome [14]." In this present study, 37 activity-related variables listed in Table 2 were used as the covariates. If the player status is 'retained,' then the dependent variable was coded as one. The other three churn status was coded as zero to make the analysis simpler. A model was developed for predicting either quit or continue to play. There were 25,000 retained cases and 75,000 cases for quitting the game. SPSS Statistics Version 22 was used to conduct the analysis. Table 7 shows statistical tests of individual predictors. "The statistical significance of individual regression coefficients (i.e., βs) was tested using the Wald chi-square statistic [15]." According to Table 8, all except 14 were significant predictors of user churn ($p < .05$). Some of the variables found to be robust separators in the boxplot analysis were insignificant (get_money, normal_chat, and party_chat). The intercept (i.e., the constant) was significant, and thus it should be included in the model. Table 9 reveals that the prediction for quitting was more accurate than for retained. The magnitude of correctly classifying user churn (97.6%) is higher than correctly predicting retained (40.3%). The overall correction prediction was 83.2%, an improvement over the chance level.

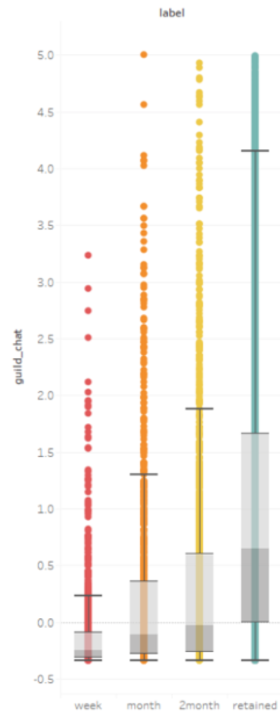

Fig. 4. *Boxplot of the number of times conducting guild chat (guild_chat) variable against four churn status*

3.3 Deep Learning

100,000 cases were recorded, out of which 70% were used for training, 15% for testing, and the remaining 15% for validation. A simple Keras sequential model was used to build a multi-layer perceptron [16]. As shown in Fig. 5, it had an input layer with 37 nodes, the first hidden layer with 32 nodes, the second hidden layer with 32 nodes, and an output layer with one node. Both the first and second hidden layers were fully connected, and 'Relu' activation was used. The output layer is also a dense layer with one neuron, and it used sigmoid activation. The optimizer function was 'stochastic gradient descent', and the loss function was 'binary crossentropy.' 'Accuracy' was tracked in addition to the loss function. The training procedure ran for 10 epochs. A 32-batch size was used to evaluate samples before updating the weights. The resulting test accuracy was 85.46%. Fig. 6 shows the model loss, and Fig. 7 represents the model accuracy for the training and validation set.

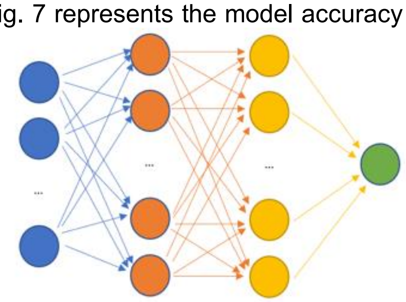

Fig. 5. *Multi-layer Perceptron Model with Two Hidden Layers [17]*

TABLE 7

BINARY LOGISTIC REGRESSION ANALYSIS OF GAME CHURN PREDICTION USING 100,000 CASES

Predictors	B	S.E.	Wald	df	Sig.	Exp(B)
cnt_dt	.170	.005	1101.693	1	.000	1.185
play_time	.150	.027	30.233	1	.000	1.162
npc_exp	.148	.034	19.127	1	.000	1.160
npc_hongmun	.009	.052	.031	1	.860	1.009
quest_exp	-.942	.053	318.998	1	.000	.390
quest_hongmun	.157	.035	20.206	1	.000	1.170
item_hongmun	.804	.043	343.010	1	.000	2.234
game_combat_time	.986	.053	351.510	1	.000	2.682
get_money	-.002	.009	.076	1	.783	.998
duel_cnt	.151	.385	.154	1	.695	1.163
duel_win	.254	.299	.721	1	.396	1.289
partybattle_cnt	.606	.088	47.248	1	.000	1.834
partybattle_win	-.362	.094	14.999	1	.000	.696
cnt_enter_inzone_solo	.819	.392	4.375	1	.036	2.269
cnt_enter_inzone_light	-4.707	.309	232.401	1	.000	.009
cnt_enter_inzone_skilled	-.152	.125	1.485	1	.223	.859
cnt_enter_inzone_normal	.555	.225	6.079	1	.014	1.742
cnt_enter_raid	.331	.066	24.880	1	.000	1.392
cnt_enter_raid_light	.066	.273	.059	1	.809	1.068
cnt_enter_bam	.067	.039	2.954	1	.086	1.069
cnt_clear_inzone_solo	-.791	.392	4.073	1	.044	.453
cnt_clear_inzone_light	4.152	.401	107.445	1	.000	63.547
cnt_clear_inzone_skilled	.122	.128	.912	1	.340	1.130
cnt_clear_inzone_normal	-1.130	.353	10.258	1	.001	.323
cnt_clear_raid	-.017	.069	.057	1	.811	.984
cnt_clear_raid_light	1.011	.271	13.910	1	.000	2.749
cnt_clear_bam	.020	.046	.192	1	.661	1.020
normal_chat	.039	.110	.122	1	.727	1.039
whisper_chat	.076	.028	7.617	1	.006	1.079
district_chat	.445	.216	4.232	1	.040	1.561
party_chat	-.059	.030	3.811	1	.051	.943
guild_chat	.269	.036	55.282	1	.000	1.309
faction_chat	.371	.134	7.682	1	.006	1.449
cnt_use_buffitem	-.558	.038	218.053	1	.000	.572
gathering_cnt	.000	.009	.000	1	.988	1.000
making_cnt	-.259	.026	102.962	1	.000	.772
constant	-1.492	.027	2969.747	1	.000	.225

Table 8: The Observed and the Predicted Frequencies for User Churn by Logistic Regression With the Cutoff of 0.50

Observed		Predicted		
		label_retained		Percentage Correct
		0	1	
label_retained	0	73164	1846	97.6
	1	14930	10070	
Overall Percentage				

4 CONCLUSIONS

The boxplot analysis revealed that players who actively engaged in battles or duels but had a few social interactions were more likely to quit the game after a while. However, the logistic regression analysis revealed that the players are likely to stay in the game if they are engaged in the battles for a more extended period. Also, not all chat types affected user churn. The temporary nature of the party chat was not a significant factor. But more permanent chat types such as guild chat and faction chat increased the chance of players not quit the game. The ones who made more things, such as arrows or pots, than others were likely to abandon the game.

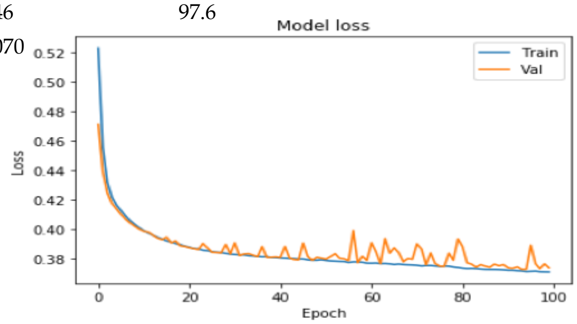

Fig. 6. *Plot of model loss for the training and validation set*

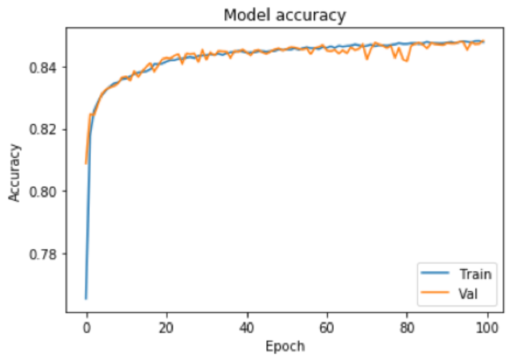

Fig. 7. *Plot of model accuracy for the training and validation set*

The ones who won more party battles left the game earliest. The players who do not have much money to spend on the game and the ones who get bored easily might contribute more to user churn. Be A simple multi-layer perceptron model outperformed the binary logistic regression but not much in terms of accurately predicting the players who might quit the game. In the future, more complex deep learning model will be studied to improve accuracy.

REFERENCES

[1] T. Dobrilova, "How Much Is the Gaming Industry Worth?" techjury, https://techjury.net/stats-about/gaming-industry-worth/. 2019.
[2] J. Kawale, A. Pal, and J. Srivastava, "Churn Prediction in MMORPGs: A Social Influence Based an Approach," Proc. IEEE International Conference Computational Science and Engineering (IEEE CSE'09), pp. 423-428, Aug. 2009, doi:10.1109/CSE.2009.80.
[3] K. Park and M. Cha, "Churn Analysis of Maximum Level Users in Online Games," Journal of Korean Institute of Information Scientists and Engineers, vol. 44, no. 3, pp. 314-322, Mar. 2017.
[4] C. Chambers, W. Feng, S. Sahu, D. Saha, and D. Brandt, " Characterizing Online Games," IEEE/ACM Transactions on Networking, vol. 18, no. 3, pp. 899-910, June 2010, doi: 10.1109/TNET.2009.2034371.
[5] NCsoft, "Rise, Strike, Agenge," NCsoft, https://www.bladeandsoul.com/uk/what-is-blade-and-soul/. 2019.
[6] Z. Bobora, J. Srivastava, K.W. Hsu, and D. Williams, "Churn prediction in mmorpgs using player motivation theories and an ensemble approach," Proc IEEE International Conference on Privacy, Security, Risk, and Trust and and IEEE International Conference on Social Computing (PASSAT/SocialCom 2011), pp. 157-164, Oct. 2011, doi: 10.1109/PASSAT/SocialCom.2011.122.
[7] N. Ducheneaut, N. Yee, E. Nickell, and R.J. Moore, "Alone together?: exploring the social dynamics of massively multiplayer online games," Proc SIGCHI Conference on Human Factors in Computing Systems (CHI '06), pp. 407-416, Apr. 2006, doi: 10.1145/1124772.1124834.
[8] H. Verhagen and M. Johansson, "Demystifying guilds: MMORPG-playing and norms," Proc. 2009 DiGRA International Conference: Breaking New Ground: Innovation in Games, Play, pp. 1-10, Sept. 2009.
[9] M. Milosevic, N. Zivic, and I. Andjelkovic, "Early churn prediction with personalized targeting in mobile social games," Journal of Korean Institute of Information Scientists and Engineers, vol. 83, pp. 326-332, Oct. 2017.
[10] MMORPG.com, "Dungeon Party" MMORPG.com, https://www.mmorpg.com/dungeon-party.
[11] T.V. Wilson, "How MMORPGs Work" howstuffworks, https://electronics.howstuffworks.com/mmorpg4.htm. 2019.
[12] Statistics How to, "Z-Score: Definition, Formula and Calculation" Statistics How to, https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/z-score/. 2019.
[13] Tableau, "Tableau Public" Tableau, https://public.tableau.com/en-us/. 2019.
[14] T. Edgar, "Exploratory Study," Research Methods for Cyber Security, T. Edgar and D. Manz, eds., Amsterdam: Elsevier, pp. 95-140, 2017.
[15] C. Peng, K. Lee, and G. Ingersoll, "An Introduction to Logistic Regression Analysis and Reporting," Journal of Educational Research, vol. 96, pp. 3-14, Apr. 2010.
[16] J. Brownlee, "How To Build Multi-Layer Perceptron Neural Network Models with Keras" Machine Learning Mastery, https://machinelearningmastery.com/build-multi-layer-perceptron-neural-network-models-keras/. 2016.
[17] J. En, "How to build your first Neural Network to predict house prices with Keras" freeCodeCamp, https://www.freecodecamp.org/news/how-to-build-your-first-neural-network-to-predict-house-prices-with-keras-f8db83049159/. 2019.