Presented at the DLSU Research Congress 2016
De La Salle University, Manila, Philippines
March 7-9, 2016

Predicting Daily Active Users for Match-3 Mobile Games

Neil Patrick Del Gallego, Oliver Bob Urag, Remedios de Dios Bulos
De La Salle University
*neil_delgallego@dlsu.edu.ph

Abstract:

One of the key measures for determining the success of a mobile game is by using Daily Active Users (DAU) as a metric. DAU is the total amount of unique users who spent considerable time in the game given a certain date. It refers to the "stickiness" of the application. A high value for DAU indicates that there is much activity and demand for the mobile game. Thus, it is essential for game development companies that develop mobile games which follows a free-to-play model, to keep the DAU value as high as possible.

We present in this paper a technique that attempts to predict the amount of daily active users for a match-3 mobile game by considering several factors that drives a game's social virality. This study is one of the few instances that use commercial game dataset for academic research. Given a certain day, we seek to predict the amount of daily active users 7 days ahead (DAU-Day7). In our prediction, we process the data extracted from the performance of two match-3 mobile games by Playlab Inc., Jungle Cubes and Dragon Cubes

A method for selection of relevant features is also discussed. We therefore attempt to use the prediction model on actual unseen data and compare the results from the outcome of our cross-validation procedure. Based from the results, M5Base algorithm performed best for both games with Jungle Cubes having significantly low error rate (using RRSE) when applied to real world test set. We discovered from our experiments that marketing expenses have high correlation with DAU-Day7 when session count and session length also have high correlation with DAU-Day7. Our results and findings will be beneficial for game companies that strictly comply with the F2P model.

Keywords: Daily Active Users; Social Virality; Mobile Games; User Retention; Predictive Analytics

1. Introduction

One of the key measures for determining the success of a mobile game is by using Daily Active Users (DAU) as a metric. DAU is the total amount of unique users who spent considerable time in the game given a certain date. It refers to the "stickiness" of the application. A high value for DAU indicates that there is much activity and demand for the mobile game. Thus, it is essential for game development companies that develop mobile games which follows a free-to-play model, to keep the DAU value as high as possible. This paper attempts to predict DAU value for two commercial match-3 mobile games, Jungle Cubes (JNC) and Dragon Cubes (DNC) that are currently owned by Playlab Inc.

Companies that follow a free-to-play business model use various analytics tools to track user's behavior and events triggered in their applications. In this study, Playlab Inc., uses Flurry Analytics and Google Play Developer Console to collect user data. One of the attributes capable of being tracked is the DAU.

In this study, we present a hypothesis that the DAU value is driven by particular attributes or events that cause a user to use the application (or in this case, play for a considerable time) extensively.

2. Dataset and Information About Attributes

Table 1 shows the overview of the datasets used for the study. Both datasets have been compiled and retrieved from Flurry Analytics[1], a commercial analytics tool used by Playlab Inc. to track user's behavior on their commercial mobile games. Some attributes were retrieved from Google Play Developer Console[2] such as user's ratings and daily crash reports. Both datasets are restricted to Android platforms only

Table 1: Overview of dataset used

Game Title	Total Downloads	Overall Rating	Timeline of Dataset
 Dragon Cubes	50,000 – 100,000	4.2 out of 5	January 05, 2015 – September 11, 2015
 Jungle Cubes	100,000 – 500,000	4.3 out of 5	January 05, 2015 – September 11, 2015

According to (Drachen, Thurau, Sifa, & Bauckhage, 2013), the dataset becomes significant if there are naturally high amounts of user activity. In our case, we extracted our dataset over the period of time specified in **Table 1** wherein there is significant development going on, many users are being acquired, and significant marketing expenses are being made to promote the game.

3. Attribute Information

This section discusses the attributes used for this study.

Table 2 contains the selected attributes influenced by the study in (Hadiji, et al., August 26-29, 2014). Attributes that are related to each other are grouped into categories.

In the Retention category, we extracted the total amount of users (Cohort Size) who installed the application on a given install date. We also extracted the retention rate (Day X retention), which determines the percentage of returning users. In the Session category, we extracted the total amount of sessions accumulated on a given date (SessionCount) and DAU value (ActiveUsers) on a given date, and the DAU value 7 days ahead (ActiveUsersDay7). Application reports from Google Play Developer Console are also extracted, namely the total number of crashes occurring in the game on a daily basis (CrashANRDay1), and the user rating (DailyAverageRating). In the Acquisition category, we considered the total amount of marketing expenses (MKTExpenses) in USD, spent to advertise the game. A high marketing expense means more advertising channels have been used to target more potential users.

Table 2: Attributes initially selected

Retention	**Usage**
Install Date Cohort Size Day X Retention	SessionCount ActiveUsers **ActiveUsersDay7**
App Reports	**Acquisition**
CrashANRDay1 DailyAverageRating	MKTExpenses
Game-specific events	
LevelPlayedEvents LevelSuccessEvents LevelFailedEvents	

Game-specific events are also tracked and these include: Accumulated number of levels played (LevelPlayedEvents), number of times a player has finished a level (LevelSuccessEvents), and the number of times a player has failed a level (LevelFailedEvents).

4. Methodology

This section contains the methodology we adopted to predict the value **DAU-Day7** from two datasets, JNC and DNC.

Using the dataset we have extracted from JNC and DNC, we explored feature selection methods and some machine learning techniques via WEKA, which are; M5Base, Decision-tree induction, Multilayer Perceptron and REPTree, to predict **DAU-Day7.** In this paper, we will only discuss M5Base and Multilayer Perception, which performed better than the rest of the algorithms.

[1] Flurry Analytics is a mobile analytics tool to analyze consumer behaviour through data observations. Website: http://www.flurry.com/solutions/analytics

[2] Google Play Developer Console is a management tool used by mobile developers for their applications released on the market. Website: https://play.google.com/apps/publish

4.1. Training and Testing Procedure

In training our models, we applied cross-validation testing on the two datasets discussed in **Table 1**. When we reached acceptable results on our prediction models, we applied it on a test set that resembles actual unseen data to see if the prediction models are good for practical use.

The test set was extracted in the following manner: Since the analytics tool continuously tracks succeeding events from users, we identified a cut-off date for our training dataset and succeeding records from both games were used as test set. Our test dataset timeline continues from our training dataset, which is between September 12, 2015 and November 9, 2015.

4.2. Feature Selection and Prediction

We formulated a feature selection method for both datasets and continuously refined our predictions models until the results were acceptable for actual deployment. The prediction model is then used for the unseen test set and we evaluated if the supposed prediction model is acceptable for future use. We discuss this further in the experiments section.

5. Experiments and Results

This section discusses how we performed our experiments and the results. We consider the following metrics for evaluation; correlation coefficient, mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE).

5.1. Correlation Analysis

We measured correlation of attributes with DAU-Day7 to determine which attributes should be included for training different prediction models. This was our first basis for feature selection. We primarily used correlation analysis to determine if there are strong relationships[3] with DAU-Day7 for both games. As seen in Table 3**,** we have observed that in JNC, strong relationships were evident between DAU-Day 7 and most of the attributes. In DNC, such strong correlation was not apparent.

JNC attributes yielded significantly higher positive correlation against **DAU-Day7**. Considering the performance of both games in the market, DNC did not have as many active users compared to JNC which may explain the low coefficient results. We therefore see how this impacts our prediction models as discussed in the next sections.

[3] We set our own threshold for classifying strong relationship to 0.7.

Table 3: Correlation coefficient on features against DAU-Day7.

	Jungle Cubes	Dragon Cubes
Cohort Size	0.6802	0.4655
Day 1 Retention	-0.0054	-0.2016
Sessions	**0.8241**	0.4261
ActiveUsers	**0.9037**	0.4630
LevelPlayedEvents	0.6817	0.4226
LevelSuccessEvents	**0.7279**	0.4700
LevelFailedEvents	0.6217	0.3111
MKTExpenses	**0.7379**	0.3747
TotalPurchases	**0.8507**	0.1577
AverageRating	0.0803	0.1578
CrashesANRDay1	0.6194	0.2668
AvgSessionSeconds	-0.0595	-0.2073
MedianSessionSeconds	0.1644	-0.0938

5.2. Feature Selection

We filtered out the unneeded attributes primarily influenced by using established feature selection techniques. We use a wrapper scheme for feature selection for learning algorithms. Prior to creating a prediction model, the wrapper scheme selects features that have the highest merit, given a machine learning algorithm. This means that features selected are deemed relevant on a training algorithm, which supposedly will increase the model's accuracy significantly. This technique is discussed in (Kohavi & John, 1997).

Basing from the initial features "proposed" to be selected by using this scheme, we manually selected, or removed some features deemed significant by this technique. Solely relying on the wrapper scheme still induced noise on the final outcome of the model. We based our manual selection method through correlation analysis and repeated observations on how it affects the overall accuracy of the model.

5.3. Prediction

This section discusses the prediction models and their outcomes. We shall only discuss M5Base and multilayer perceptron in this section as they yield better results.

ISSN 2449-3309

5.3.1.M5Base

The M5Base algorithm from (Wang & Witten, 1997) performed best for both games. The prediction model is also practical and easy to use due to its nature of inducing decision trees with regression formulas.

Table 4: Selected features for M5Base

Jungle Cubes	Dragon Cubes
Cohort Size	CrashesANRDay1
Day 1 Retention	LevelFailedEvents
LevelFailedEvents	LevelSuccessEvents
LevelSuccessEvents	Sessions
Sessions	MKTExpenses
MKTExpenses	AvgSessionSeconds
TotalPurchases	
MedianSessionSeconds	

Using the feature selection method discussed in the previous section, the features shown in **Table 4** were used for M5Base.

Figure 1 shows the correlation coefficient results using M5Base. Notice that JNC has high correlation. DNC has high correlation on its training set but using the test set, the correlation has dropped significantly which makes the model questionable for real world use.

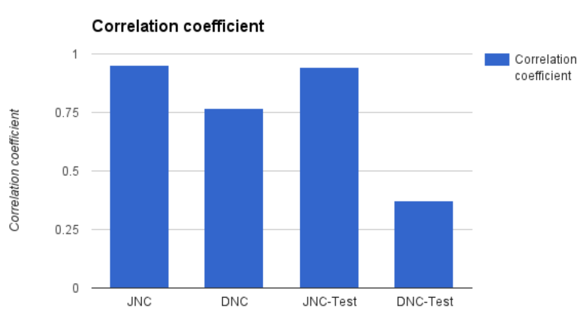

Figure 1: Correlation coefficient results in M5Base

Figure 2 shows the error tendencies. We can see that the MAE and RMSE of JNC are somewhat higher than that of DNC. However, looking into RAE and RRSE, we see that JNC outperforms DNC by a huge margin. The prediction model becomes more reliable than a simple predictor (averaging) for JNC's case.

5.3.2.Multilayer Perceptron

We attempted to use artificial neural network using Multilayer Perceptron as learning algorithm for predicting **DAU-Day7** on the impression that this learning algorithm is flexible and has various parameters involved that we can continuously adjust to improve the accuracy of the model.

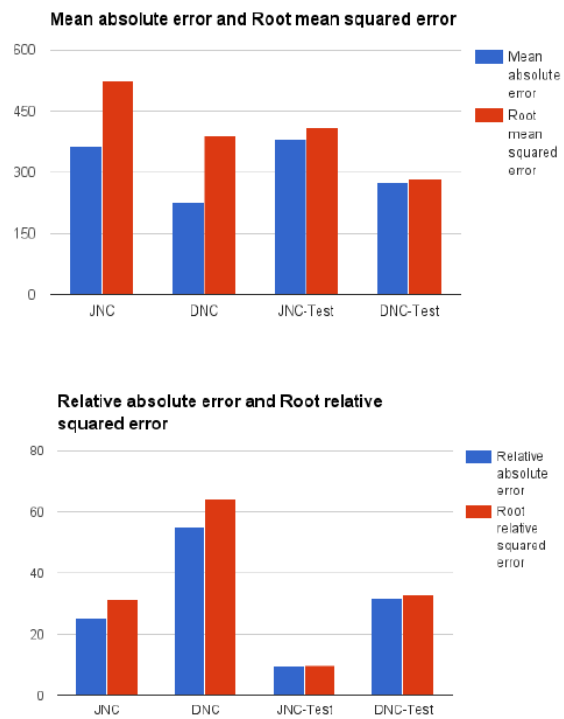

Figure 2: Error tendencies for m5Base

Table 5 shows the selected features for multilayer perceptron using our method specified in the previous section.

Table 5: Selected features for Multilayer Perceptron

Jungle Cubes	Dragon Cubes
Cohort Size	Cohort Size
MKTExpenses	MKTExpenses
TotalPurchases	
MedianSessionSeconds	
AvgSessionSeconds	

We found that the number of hidden layers appropriate for this experiment is the total number of attributes and classes of our dataset. We relied on acceptable training time to modify the learning rate

ISSN 2449-3309

and the total number of epochs[4]. Using the default values provided did not yield acceptable results.

Referring to **Figure 4**, it can be observed that multilayer perceptron performed acceptably for JNC while not for DNC. The error measures for DNC for multilayer perceptron are comparatively higher than that of M5Base.

Applying multilayer perceptron on our test set, we observed that JNC have a very low correlation coefficient as shown in **Figure 3,** which is very different from the outcome of M5Base. We noticed that the MAE and RMSE have significantly increased which makes multilayer perceptron less than ideal for practical use in JNC.

Using multilayer perceptron for DNC, the results from the test set closely resemble results from the cross-validation tests in M5Base. The RAE and RRSE is more than 50% which may indicate that using multilayer perceptron is only halfway better than a simple predictor. Furthermore, only two features were selected from our method which makes the model questionable for practical use.

6. Observations and Recommendations

M5Base predicts better in JNC than DNC From a business point-of-view, JNC is more profitable in the market than DNC. This may explain why the prediction model of JNC is better than DNC as it contains more reliable data. There is significantly more amount of daily users in JNC than DNC.

We observe that JNC yields better results than DNC as there more features in JNC that have a strong positive correlation with DAU-Day7 as discussed in our correlation analysis section, which contributed to a better prediction model for JNC.

M5Base yields the highest correlation coefficient across our JNC tests and also produced low error rates. Using the model for practical use, we see that while there is little difference in MAE and RMSE, the RAE and RRSE of JNC yields only 9%. This makes it suitable for real world use.

7. Conclusion and Future Work

We attempted to verify if it is possible to predict **DAU-Day7**.

[4] Actual count from best results has 6 hidden layers. Learning rate is set to 0.001, number of epochs set to 5000.

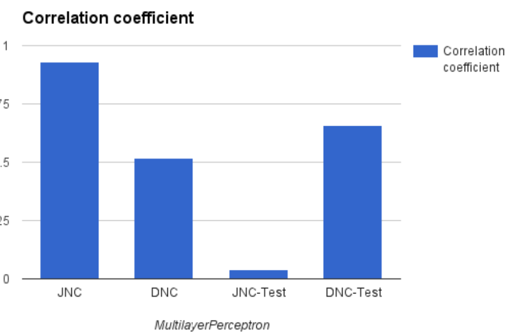

Figure 3: Correlation coefficient in Multilayer Perceptron

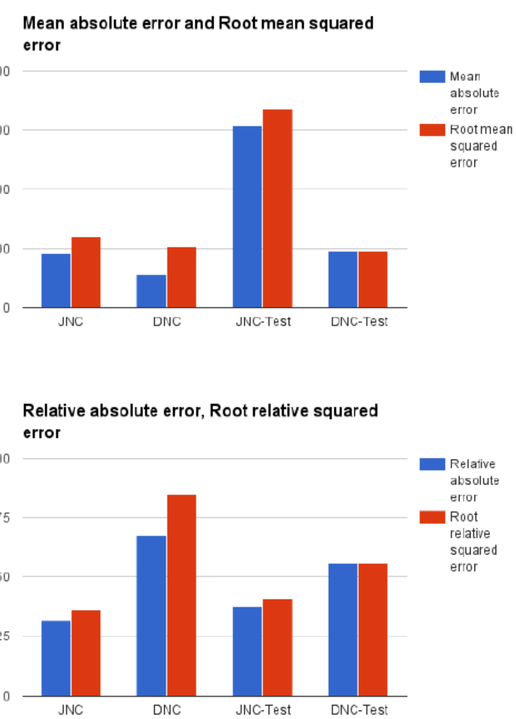

Figure 4: Error tendencies for Multilayer Perceptron

We considered how features are frequently selected by using our proposed feature selection scheme. We observed from our experiments, those that relate to session activity (total number of sessions and session length) were selected on all cases. This supported the findings of (Hadiji, et al., August 26-29, 2014) wherein session count and length has an effect on application virality.

7.1. Marketing expenses, session count and session length

We observed that the feature, **MKTExpenses,** was selected on all cases. **MKTExpenses** for JNC has high correlation with **DAU-Day7** (0.73). While **MKTExpenses** is also being selected as a contributing feature for DNC, it yields a low correlation value. This gives support to Playlab's situation during promotion of DNC. User acquisition campaigns acquired enough users to install the game on their devices but majority of the users left the game before Day 7. This could mean that users are not engaged to the game which motivated them to leave. This explains the low correlation of these two features, which is contradicting to the result on JNC. Based from our study, we propose a finding that **MKTExpenses** gets high correlation with **DAU-Day7** once the game has enough enjoyable content to keep users engaged. To make engagement factors high, business decisions should improve the outcome of session length and number of sessions (**Sessions** and **MedianSessionSeconds** should have strong correlation with **DAU-Day7**). Once those are set, increasing marketing expenses for more user acquisition campaigns will also increase the potential of gaining additional daily active users.

7.2. Factors in game design

We observed that JNC performed better and has a better accuracy on M5Base and Multilayer Perceptron unlike DNC. We therefore conclude that the success of a game contributes to better quality of data and discovery of more patterns than if a game did not really do well in the market. We deduced that DNC, which is less commercially successful, has underlying problems in its game design. None of its features yield high correlation value with **DAU-Day7**. Events and patterns from the data were not observable.

We propose as future work that games like DNC should also have a concrete model to quantify the engagement factor of the game. Inferring the fun factor of a game may also be modelled by gathering user sentiment or initial feedback. Relying on events that only consider session length, triggered events and marketing expenses may not fully capture such cases like this.

7.3. Considering the quality of the game

We observed that we yield a somewhat high positive correlation value between **CrashesANRDay1** and **DAU-Day7**. This is misleading because crashes are labelled as negative events which may motivate users to leave the game. Such negative events were not properly modelled in our study. It would be beneficial if one would consider the impact of a game that is defective. We propose that various software quality metrics be measured as well.

To conclude our study, **DAU-Day7** can be predicted using our proposed method for successful games. JNC yields the lowest error rate which makes it ideal for M5Base proposed in this paper. On the case of DNC, we conclude that it does not perform as expected and may not be practical to use such model. We propose that in games wherein cases are similar to DNC, one should consider taking into consideration factors in game design and quality of the game. We recommend using M5Base as the model to be used for practical application due to how well it covered different scenario in the data.

8. References

Drachen, Thurau, Sifa, & Bauckhage. (2013). A comparison of methods for player clustering via behavioral telemetry. *International Conference on the Foundations of Digital Games*, (pp. 245-252). Greece.

Hadiji, Sifa, Drachen, Thurau, Kersting, & Bauckhage. (August 26-29, 2014). Predicting player churn in the wild. *2014 {IEEE} Conference on Computational Intelligence and Games, {CIG}*, (pp. 1-8). Dortmund, Germany.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence: Special issues on relevance*, (pp. 273-324).

Wang, Y., & Witten, I. H. (1997). Induction of model trees for predicting continuous classes. *Poster papers of the 9th European Conference on Machine Learning.* Springer.

ISSN 2449-3309